Units in Speech Perception*

Zinny Sang Bond

*Sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

ACKNOWLEDGMENTS

The help of many people has made this study possible. Above all, I want to thank my adviser, Professor Ilse Lehiste, whose insistence on clear formulation of ideas prevented me from trying many unworkable schemes. I also want to thank the members of my dissertation committee: Professors Catherine Callaghan, Gaberell Drachman, Arnold Zwicky, and, particularly, Neil Johnson. Preston Carmichael, technician, spent many hours helping me design and assemble instrumentation. I sincerely appreciate his help. Tom Whitney gave much assistance in devising and using computer programs for data processing. I thank him for his kind assistance. I also want to thank the Ohio State University Instruction and Research Computer Center for permitting me to use their facilities. I am grateful to the "Phonetics Research Group"--Sara Garnes, Dick Gregorski, Linda Shockey, and Mary Wendell--for listening patiently to my problems and offering many helpful suggestions. I appreciate the help of all the members of the Ohio State University Linguistics Department who kindly served as subjects in many of these experiments. Finally, I want to thank my family for their patience and moral support while I was writing this dissertation.

This work was sponsored in part by the National Science Foundation through Grant GN-534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	v
LIST OF FIGURES	viii
INTRODUCTION	1
Chapter	
I. MODELS OF SPEECH PERCEPTION	4
Behaviorism Information Theory Filtering The Motor Theory of Speech Perception Analysis by Synthesis Perceptual Strategies	
II. THE PERCEPTION OF SUB-PHONEMIC DIFFERENCES	25
Method Results Discussion	
III. THE PERCEPTION OF OBSTRUENT CLUSTERS	45
Method Results Discussion	
IV. SYNTACTIC UNITS IN PERCEPTION	76
Method Results Discussion	
V. CONCLUSION	91
The Need for Perceptual Units Implications for Perception Models	
BIBLIOGRAPHY	98

LIST OF TABLES

Table		Page
1.	Per Cent Correct Identifications	30
2.	Consistency of Subjects' Responses	32
3.	Subjects' Performance in Relation to Judgments of Ease and Difficulty	33
4.	Ease and Difficulty of Word Pairs	34
5.	Reaction Time for Subjects with Training in Phonetics for the Pairs Wade/weighed, etc.	35
6.	Reaction Time for Subjects with Training in Phonetics for the Pairs Baste/based, etc.	36
7.	Reaction Time for Phonetically Untrained Subjects for the Pairs Wade/weighed, etc.	37
8.	Reaction Time for Phonetically Untrained Subjects for the Pairs Baste/based, etc.	38
9.	Reaction Time to Productions Labeled Consistently. . . .	39
10.	Reaction Time to the Mono-morphemic and Bi-morphemic Words Weighed/wade, etc.	41
11.	Reaction Time to the Mono-morphemic and Bi-morphemic Words Lacks/lax, etc.	42
12.	Reaction Time to the Mono-morphemic and Bi-morphemic Words Missed/mist, etc.	43
13.	All Responses--Signal to Noise Ratio: +12 d.b.	53
14.	All Responses--Signal to Noise Ratio: 0 d.b.	53
15.	All Responses--Signal to Noise Ratio: -6 d.b.	54
16.	All Written Responses--Signal to Noise Ratio: +12 d.b. .	55
17.	All Written Responses--Signal to Noise Ratio: 0 d.b. .	55
18.	All Written Responses--Signal to Noise Ratio: -6 d.b. .	56

Table | Page

19. Total Spoken Responses--Signal to Noise Ratio: +12 d.b. . 57

20. Total Spoken Responses--Signal to Noise Ratio: 0 d.b. . 57

21. Total Spoken Responses--Signal to Noise Ratio: -6 d.b. . 58

22. Written Responses for Two-Syllable Words--Signal to Noise Ratio: +12 d.b. 59

23. Written Responses for Two-Syllable Words--Signal to Noise Ratio: 0 d.b. 59

24. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: +12 d.b. 60

25. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: 0 d.b. 60

26. Written Responses for Two-Syllable Words--Signal to Noise Ratio: -6 d.b. 61

27. Spoken Responses for Two-Syllable Words--Signal to Noise Ratio: -6 d.b. 62

28. Spoken Responses for [ɪ]--Signal to Noise Ratio: +12 d.b. 63

29. Spoken Responses for [æ]--Signal to Noise Ratio: +12 d.b. 63

30. Spoken Responses for [u] and [ou]--Signal to Noise Ratio: +12 d.b. 63

31. Spoken Responses for [ɪ]--Signal to Noise Ratio: 0 d.b. . 64

32. Spoken Responses for [æ]--Signal to Noise Ratio: 0 d.b. . 64

33. Spoken Responses for [u] and [ou]--Signal to Noise Ratio: 0 d.b. 64

34. Spoken Responses for [ɪ]--Signal to Noise Ratio: -6 d.b. 65

35. Spoken Responses for [æ]--Signal to Noise Ratio: -6 d.b. 65

36. Spoken Responses for [u] and [ou]--Signal to Noise Ratio: -6 d.b. 65

37. Written Responses for [æ]--Signal to Noise Ratio: -6 d.b. 66

38. Written Responses for [ɪ]--Signal to Noise Ratio: -6 d.b. 66

39. Written Responses for [u] and [ou]--Signal to Noise Ratio: -6 d.b. 66

Table		Page
40.	Written Responses for [I]--Signal to Noise Ratio: 0 d.b.	67
41.	Written Responses for [æ]--Signal to Noise Ratio: 0 d.b.	67
42.	Written Responses for [ʊ] and [oʊ]--Signal to Noise Ratio: 0 d.b.	67
43.	Written Responses for [æ]--Signal to Noise Ratio: +12 d.b.	68
44.	Written Responses for [ɪ]--Signal to Noise Ratio: +12 d.b.	68
45.	Written Responses for [u] and [oʊ]--Signal to Noise Ratio: +12 d.b.	68
46.	Written Responses for Bi-morphemic Words--Signal to Noise Ratio: +12 d.b.	70
47.	Written Responses for Bi-morphemic Words--Signal to Noise Ratio: 0 d.b.	70
48.	Written Responses for Bi-morphemic Words--Signal to Noise Ratio: -6 d.b.	70
49.	Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: +12 d.b.	71
50.	Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: 0 d.b.	71
51.	Spoken Responses for Bi-morphemic Words--Signal to Noise Ratio: -6 d.b.	71
52.	Reaction Time for Correct and Incorrect Responses	72
53.	Reaction Time to Consonant Clusters	73
54.	Mean Reaction Time to Clicks	81
55.	Click Localization: Per Cent Correct	88

LIST OF FIGURES

Figure		Page
1.	Three-stage Mediation-integration Model	8
2.	Model of a Closed Cycle Control System for Speaking . . .	13
3.	A Model of Speech Communication	14
4.	Analysis by Synthesis Model	20
5.	Model for the Speech-generating and Speech-perception Process	21
6.	Instrumentation for Experiment Testing the Perception of Sub-phonemic Phonetic Differences	28
7.	Per Cent Correct Identifications for Each Word Pair . . .	31
8.	Instrumentation for Adding Noise to Stimulus Tape . . .	50
9.	Instrumentation for "Click" Experiment	80
10.	Reaction Time to Clicks in Consonants Preceding Stressed Vowels and to Clicks in Consonants Preceding Unstressed Vowels	82
11.	Reaction Time to Clicks in Stressed Vowels and in Unstressed Vowels	83
12.	Simple Reaction Time to Click, and Reaction Time to Click in a Constituent Boundary	84
13.	Click Localization When the Click Occurs in a Constituent Boundary	85
14.	Click Localization	85-87
15.	Click Localization in Stressed and Unstressed Vowels . .	88

INTRODUCTION

Speech perception, as a field of empirical investigation, is very much involved with linguistics: a model of speech perception is crucially dependent on a model of language, since the model of language tells the perception theorist what it is that the listener has to perceive.

Thus, historically, there has been a tendency for models of speech perception to be related to the current linguistic models of language. The early models of speech perception are not specific enough, by current standards, simply because the model of language that the theorist was dealing with was not a very complex model--language was conceived to be something like a series of words strung together.

As more complicated and more precise linguistic models become current, the theorizing about speech perception also became more precise and more experimentally oriented. Thus, structural linguistics of the 1940's and 1950's led to experimental work which assumed that the phoneme, or some unit very much like a phoneme, was the perceptual unit in phonology. The problem in understanding speech perception was then seen as discovering how a listener can 'translate' or 'decode' a continuous acoustic signal into discrete phonemes. And, though alternative suggestions have been made, most theorists still assume that the incoming speech signal is represented in some phoneme-like units as the first step in speech perception.

Experimental work on higher-level perceptual units, related to the syntactic structure of a sentence, has begun quite recently. Some early theorists have advanced ideas of what is involved in understanding sentences, but, again, the work could not lead to any precise theoretical formulations until a fairly adequate theory of syntax became available; thus, almost all empirical studies involving the perception of syntactic units assume that the syntactic relationships described in transformational grammar are involved in speech perception at some level. However, the experiments have tended not to separate perceptual effects from memory effects; and there is no agreement--such as implicitly exists in theories of the perception of phonological segments--whether there are some syntactic units involved in perception and, if so, what these units are.

Generative phonology, which does not assume any unit equivalent to the traditional phoneme, has not so far led to any experimental work on speech perception, though it is intimately related to models of speech perception involving analysis-by-synthesis.

In this study, the attempt is made to examine some units that function in speech perception. The first chapter contains a survey of models that have been proposed to account for speech perception. The survey includes some models because of the historical background they provide, even though the models make no specific predictions about units in speech perception. More recent models make certain predictions about perceptual units, and these will be pointed out when the theoretical implications of the perceptual models are discussed.

Three experiments are reported. The first experiment involves a subject's ability to make use of sub-phonemic phonetic differences.

Subjects are asked to identify productions of mono-morphemic and bi-morphemic words of identical phonemic shape, e.g., lax vs. lacks. The purpose of the experiment is two-fold: to determine what a 'baseline' for perception is--what is the least amount of phonetic difference that can be used for linguistic purposes--and to determine if the traditional phoneme, which is often accepted as the perceptual unit, defines a lower limit below which a listener can not make use of phonetic differences.

The second experiment involves the perception of obstruent clusters. Subjects are asked to identify words with reversible obstruent clusters, such as task vs. tax, in the presence of noise. The purpose of the experiment is to determine whether consonant clusters are coded 'phoneme-by-phoneme', as the traditional assumptions would imply, or if subjects employ some alternative perceptual mechanisms.

The third experiment seeks to determine perceptual units in syntax. Subjects are asked to respond, by pressing a button, when they hear a 'click' in a sentence. From reaction time to the click, the effects of a phonologically defined phrase on perceptual segmentation can be determined.

Finally, the implications of the experimental studies to models of speech perception are discussed.

CHAPTER ONE

MODELS OF SPEECH PERCEPTION

The purpose of this chapter is to provide some historical background and to present the current ideas of theorists attempting to account for speech perception. Not all of the models that will be discussed in this chapter make specific predictions about what units are involved in speech perception, but they are included simply because many are interesting in themselves or for historical reasons.

No attempt will be made to evaluate the adequacy of any of these models in this chapter. Rather, the models that still hold promise will be discussed in the last chapter in terms of the theoretical implications of the empirical studies reported in this work.

Models of speech perception have been classified under the following headings: behavioristic models, information theory models, motor theories, analysis by synthesis models, models proposing 'filtering' as a primary device, and models depending on perceptual strategies.

Behaviorism

There is a long behaviorist tradition of theories of speech perception. Appropriately enough, it begins with J. B. Watson (1930). Watson's general behaviorist position is well known, and his views of language--not developed in any great detail--follow from it clearly. Since he refuses to postulate any "mentalistic constructs,"

he discusses language in observable, physicalistic terms. Language is simply a "manipulative habit of the vocal tract" (Watson, p. 225). When a person learns to speak, he develops a conditioned response--some movement of the vocal tract--for every object and situation in his external environment. These conditioned responses are equivalent to words. Such internalized kinaesthetic responses can call out further reponses in the same way as the objects for which they serve as substitutes do; because of these kinaesthetic verbal substitutes, a person carries the world around with him; he can manipulate the world (think) by means of series of motor responses.

Sentences, and other language sequences, are accounted for by the following example: a child hears the bed time prayer "Now I lay me down to sleep..." The first few times he hears it, the first word of the sentence, "now," makes the child produce the motor response which is his internal equivalent of "now;" similarly "I" leads to internalized "I," etc. After repeated experiences, the motor response "now" will lead directly to the motor response "I," with no necessary intervening step. At this point, the child has learned the sentence. Spontaneous speech, Watson believes, follows essentially the same principles: some stimulus touches off old verbal organization.

Speech perception offers no particular difficulty: the incoming stimulus makes the listener form the equivalent kinaesthetic-motor responses. Watson, therefore, is postulating a simple motor theory of speech perception, involving incipient muscle activity.

In Language, Bloomfield (1933) offers a much more sophisticated analysis of language, but his outlook is essentially behavioristic.

Bloomfield analyzes an event involving speech by means of a little scene with two characters, Jack and Jill. Externally, the action is quite simple: Jack and Jill are walking along a road; Jill makes a series of noises with her vocal tract; Jack climbs a fence, and brings Jill an apple from a nearby tree.

Looking at the scene more analytically, there are a number of practical events preceding the act of speech. These practical events are quite complex, but taken together, they can be considered as a stimulus for Jill. As a speaking human, Jill has a choice: she can make a direct response (go get the apple), or she can make a linguistic substitute response (ask Jack for the apple). For Jack, the speech is a substitute linguistic stimulus, which makes him produce a particular response.

Essentially, speech enables stimuli and responses to occur in different individuals, as indicated in the following diagram:

$$S \rightarrow r \cdots\cdots s \rightarrow R$$

Bloomfield is not very specific in discussing what is involved in Jack's reception of the message. In relation to phonology, Bloomfield argues that speakers of a language habitually and conventionally discriminate some features of sound and ignore others; presumably, then, there are distinctive properties of sound to which Jack is sensitive. These encode the message.

The behaviorist tradition is carried on in the 1950's by the psychologists B. F. Skinner (1957), O. H. Mowren (1954), and C. E. Osgood (1963).

Mowrer does not offer a complete theory of language, but an analysis of declarative sentences in stimulus-response (henceforth S-R)

terminology. Essentially, he suggests that a sentence is an arrangement for conditioning the meaning reaction produced by the predicate to the stimulation aroused by the meaning reaction elicited by the subject. In other words, a subject-predicate sentence is to be considered a conditioning device.

The conditioning device operates in the following way. When the listener hears any word in his vocabulary, there is aroused in him a unique "meaning response." When he hears a sentence, for example, "Tom is a thief," first there is aroused in the listener a "meaning response" which is his internal representation of the word "Tom" as well as of the physical Tom. Then, because a sentence is a conditioning device, to this "meaning response" is added the "meaning response" of "thief." As a consequence, the listener comes to respond differently to the physical Tom; he will avoid him, perhaps, and not lend him money. In short, he will treat Tom as a thief.

One of the most thorough attempts to explain language behavior in S-R terms is B. F. Skinner's book Verbal Behavior (1958). Skinner declines to speculate about non-observable language phenomena; rather, he sees the task of the science of verbal behavior to determine the laws governing verbal behavior. These laws concern the predictability and control of particular verbal responses. That is, the task is accomplished when it is possible to predict what a person will say.

Because of this goal, and because he rejects non-observables, Skinner has little to say about internal phenomena such as perception. He does offer a few suggestions. First, Skinner defines a unit of verbal behavior as anything that is under the independent control of a manipulable (stimulus) variable. This unit can be as large as a

whole phrase, such as "How are you?", or as small as a change in fundamental frequency, used to ask a question. In order for language to function at all, these units must lead to different responses by listeners. Secondly, Skinner points out that at any time in sequential verbal behavior, e.g. sentences, what has been said before sharply limits what will be said next: there is redundancy in language. Presumably, the listener can also take advantage of such redundancy.

But Skinner does not attempt to present any theory of speech perception; the few suggestions that he makes do not detract from his basic assumption that perception can not be separated from responses in any meaningful way.

C. E. Osgood also offers a behavioristic theory of speech (Osgood, 1963), which he calls a three-stage mediation model. Unlike Skinner, Osgood is quite ready to postulate mechanisms internal to the speaker and listener. Rather than being concerned only with observable stimuli and responses, Osgood wants to fill the "black box" of the organism with intervening S-R constructs. Osgood's three-stage model is represented below.

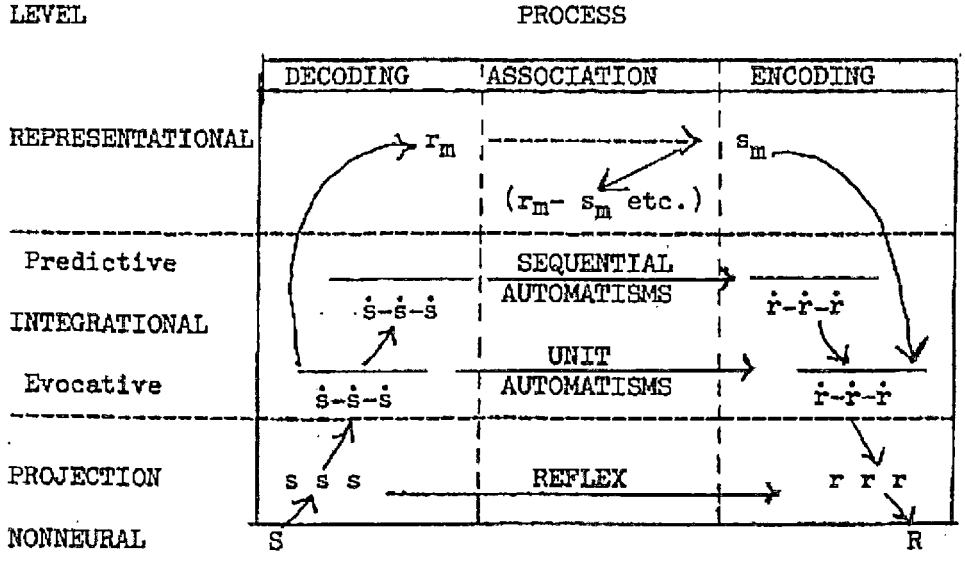


Fig. 1. Three-stage mediation-integration model.

(by permission of Charles E. Osgood)

Osgood's model differs from Skinnerian S-R models in two ways. First, Osgood postulated mediating responses (r_m). These internal r_m's are a fractional, easily differentiable, part of an original overt response. Since the original response was elicited by some stimulus, the fractional r_m becomes an internal representation of the stimulus. The internal r_m's, in turn, can lead to various instrumental acts. Essentially, Osgood hopes to account for meaning by these internal representations. These internal representations, however, are quite complex; basically, Osgood holds that words are coded by means of a simultaneous bundle of semantic features (Osgood, 1963).

Secondly, Osgood postulates stimulus integration (S-S learning) and response integration (R-R learning) to account for the perceptual and motor complexity found in speech. He argues that, in perception, the greater the frequency with which stimulus events have been paired in the input experience of the organism, the greater will be the tendency for their central neural correlates to activate each other. In other words, a partial sensory input will become adequate to trigger the whole; it will lead to what the Gestalt psychologists called "closure."

This closure principle can only operate if there are perceptual units which function as wholes. These units must meet three criteria: they must be highly redundant, they must be fairly frequent in occurrence, and they must not exceed certain temporal limits. The most likely perceptual units are words.

In perceiving a sentence, the phonetic information is adequate to trigger the phonological representation of a particular word, e.g.

play. The context of the sentence then determines the semantic interpretation of the word. Given, for example, the sentence "The play got rave reviews," the word play will be interpreted as a noun on the basis of the frame Determiner ____ verb. The word review will eliminate the interpretation of play in the sense of gambling. On the basis of such linguistic information and on the basis of non-linguistic context, the listener will arrive at the intended message.

More recently, psychologists, even though they may consider themselves behaviorists, have broken away from S-R formulations altogether.

In his very interesting book, The Senses Considered as Perceptual Systems, James J. Gibson (1966) emphasizes the information contained in stimulation, rather than the discrete responses of separate sensory systems. Therefore, he rejects the traditional decomposition of a complex sound into a combination of pitch, duration, and loudness specifications in order to describe the stimulus. He considers it a better approach to look for higher-order variables characteristic of the stimulus:

> "In meaningful sounds, these variables can be combined to yield higher-order variables of staggering complexity. But these mathematical complexities seem nevertheless to be the simplicities of auditory informaiton, and it is just these variables that are distinguished naturally by an auditory system." (p. 87).

In other words, it is a mistake to think that the perceptual system "builds up" complex stimuli from simple components; rather, complex stimuli are responded to directly.

The higher-order variables have not been studied for most types of meaningful sound, but there have been a few attempts to study

such variables in the acoustic speech signal. According to Gibson, frequency ratios and the relational patterns of frequencies are the invariants provided by the speech signal.

The pick-up of phonemes is a direct one-stage process; however, the apprehension of things referred to--a semantic decoding of the speech signal--is a two-stage process since not only the speech sounds but what they stand for have to be apprehended. "The acoustic sounds of speech specify the consonants, vowels, syllables, and words of speech; the parts of speech in turn specify something else." (p. 91).

The structure of speech can be analyzed at various levels, hierarchically organized, and each level has some unit appropriate to it: at each level, there is an appropriate stimulus unit for the perceptual system.

Information Theory

During the 1950's, information theory provided conceptual structures by which all types of communication--defined as the transmission of information--could be analyzed. Theorists concerned with speech also tried to apply the concepts of information theory to their field, and developed models of speech communication. These speech communication models discussed both a speaker and a hearer, but tended to emphasize the former. Many models of the speech communication system were proposed; these are summarized by Grant Fairbanks (1954), who also presents one of the most detailed analyses of speech from this point of view. However, most of his discussion concerns speech production. Perception is discussed almost exclusively in terms of its role in feedback: the speaker monitors his own output and changes his output

when it does not meet the criteria set by the input to the speech systems.

Fairbanks' model is reproduced in Fig. 2. Essentially, the model offers the following analysis of speech production: an input signal to the speech mechanisms results in some output; this output is compared with the stored input; if the output has not yet reached the target specified by the input, an error signal is sent out to adjust the output.

There are several interesting points concerning the speech model. First, Fairbanks postulates a "unit of control." Although he does not go into detail, he suggests that the unit of speech control is not to be identified with any currently recognized phonetic unit; rather, the unit of speech control is a "semi-periodic, relatively long, articulatory cycle" (p. 138). Secondly, the model implies that certain steady-state outputs are the goals of the speech mechanism and that transitions are only by-products. In Fairbanks' words:

> "It is to be emphasized that the steady states are the primary objectives, the targets. The transitions are useful incidents on the way to the targets. The roles of both are probably very analogous when the dynamic speech output is perceived by an independent listeners." (p. 139)

Fairbanks has little to say about speech perception directly. Presumably, perception follows the path described for feedback. Whether the message is analyzed directly or whether it is compared in the comparator with a possible message--as in motor theories of speech perception--is not specified in Fairbanks' model.

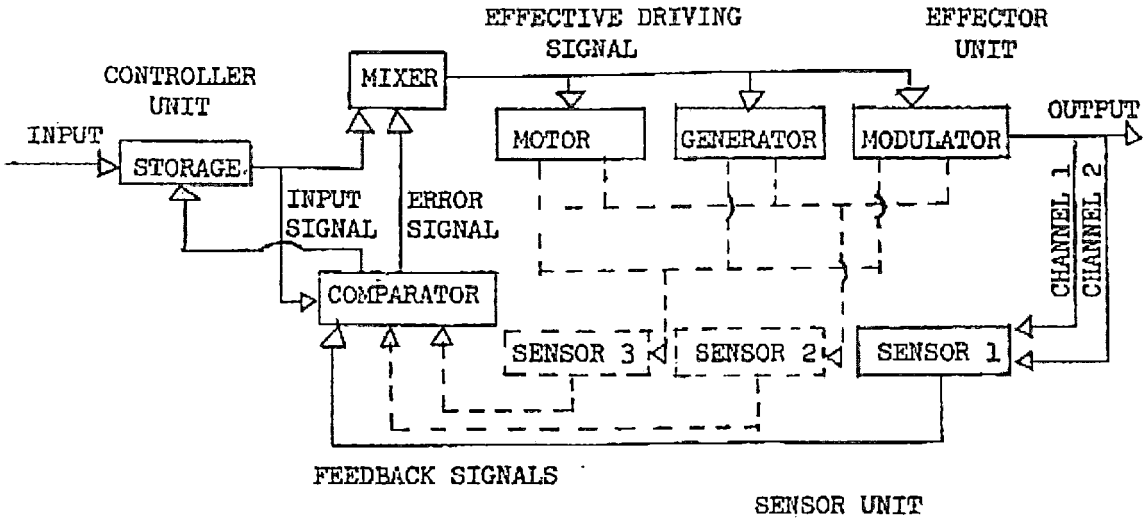


Fig. 2. Model of a closed cycle control system for speaking. (Grant Fairbanks, "A Theory of the Speech Mechanism as a Servo-System." Journal of Speech and Hearing Disorders 19 (1954). By permission of the American Speech and Hearing Association).

Although it uses concepts from information theory, Hockett's model of speech communication (1956) is much more linguistic in orientation than Fairbanks' model, at least in the sense that linguistic terminology is applied to various processes. However, Hockett cautions that the 'phoneme' and 'morpheme' of internal circuitry are not to be strictly equated with the phoneme and morpheme of linguistics.

Hockett's model (Fig. 3) represents the internal mechanisms necessary for Jill to communicate with Jack. First, a sequence of morphemes is emitted by GHQ (grammatical headquarters); then the morphemes are recoded into a discrete flow of phonemes by morphophonemic processes. Finally, the phonemes become a continuous speech signal in the "speech transmitter." The speaker monitors his own speech signal, but he does not use feedback to adjust the output continuously.

The listener uses the same communications system, but the speech

receiver sends the signal through in the other direction; the speech receiver picks up the signal and transduces it into a discrete flow of phonemes; the phonemes are assembled into morphemes and submitted to GHQ. A listener understands a message when his GHQ is going through the same "states" as the speaker's GHQ. Hockett also suggests that a listener decodes an incoming signal partly by comparing it with the articulatory motions that the listener would have to make to produce the signal.

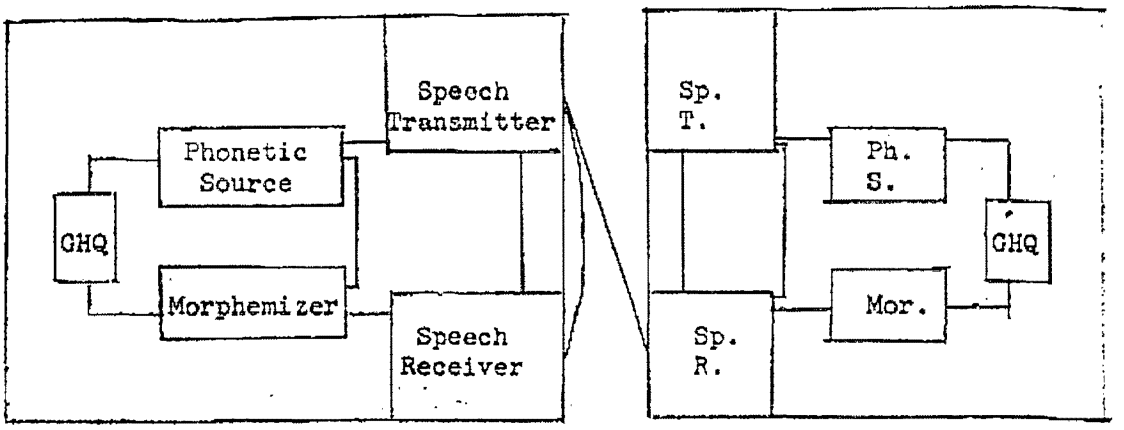


Fig. 3. A model of speech communication. (Charles Hockett, A Manual of Phonology, 1955, by permission of Indiana University Publications in Anthropology and Linguistics and Prof. Charles F. Hockett.)

Filtering

In his article "On the Process of Speech Perception," J. C. R. Licklider (1952) analyzes the process of speech perception into three main operations: translation of the speech signal into a form suitable for the nervous system, identification of speech elements, and comprehension of meaning.

The first process is performed by the cochlea; the signal is mechanically analyzed in terms of frequency and intensity in such a way that the output is somewhat similar to a sound spectrogram. However, since the frequency analysis of the cochlea is not very selective, the signal is sharpened further up the auditory pathways. Thus, the input to the perceptual mechanism consists of a sharpened frequency analysis of the acoustic signal, coded in terms of origin on the cochlea, and intensity, coded in terms of density of discharge. Furthermore, there is a representation of the fundamental frequencies of the periodic components of the acoustic signal.

The second process, identification of speech elements, could be performed by one of two mechanisms, a correlator or a filter. A correlator is essentially a device for matching the incoming signal against an internally stored representation (or a representation created by rules). A filter, on the other hand, has the required patterns built into its structure; the identification of the incoming signal is made on the basis of which filter the signal passes through most successfully. Although the choice is tentative, Licklider favors the filter model as the device which identifies speech elements.

Comprehension, on the other hand, can best be explained as an active process. Therefore, Licklider argues that comprehension of meaning involves matching the input to a set of internal patterns. Although he does not say this, Licklider would probably maintain that these patterns are generated as needed.

Licklider's model, therefore, is very much like analysis-by-synthesis for the processing of sentences. For smaller units, however, Licklider prefers the more direct analysis provided by filtering.

A "filtering" theory, differing in interesting ways from Licklid has been recently developed by Wayne A. Wickelgren (1969a, 1969b). Previous theories have assumed that, no matter how speech is processe the phoneme is the primary unit of coding in perception. Wickelgren proposes a theory in which the perception and production of speech is coded in some unit that is more closely related to the traditional allophone. He calls this theory context-sensitive coding.

> "I define a context-sensitive code for words to consist of an unordered set of symbols for every word, where each symbol restricts the choice of its left and right neighbors sufficiently to determine them uniquely out of the unordered set for any given word. In this case, the unordered set, in conjunction with the dependency rules, contains all the information necessary to reconstruct a unique ordering of the symbols for each word." (1969b, p. 86)

In speech perception, context-sensitive coding would work in t following way. Each context-sensitive allophone of the language wo have a unique internal representative. This internal representativ would be activated by some conjunction of acoustic features, occurr over a period of time as long as a few hundred milliseconds. All allophone representatives would be examining the acoustic input in parallel, but only a few would be activated in response to the inpu After the set of allophones has been determined, the word represen which is most closely associated with the set of allophones can be selected.

Wickelgren claims that his theory eliminates two of the major problems associated with perception models which postulate phoneme as the basic units: first, there is no need to segment the acousti wave form; second, it is more likely--although the evidence is not that there is invariance in the acoustic signal for allophones.

The model of speech perception proposed by L. V. Bondarko and others (Bondarko et al., 1970) is designed to account for the set of operations that transform an acoustic speech signal into a sequence of words. Each word in the output would have associated with it a set of lexical and grammatical features which would be employed in understanding the message.

The model consists of hierarchically-arranged processes. At each level, there is a perceptual procedure, decision making, and a procedure for assigning a certain reliability to the decision. If no decision can be made with a threshold degree of reliability, the level outputs several possible interpretations of the input signal, and the final decision is postponed. The final decision may not be made, in fact, until the last stage--the recognition of the meaning of the utterance.

The first stage of the perceptual process is auditory analysis. The output of the cochlea is described in the set of parameters that are relevant in the perception of speech. The output of the auditory analysis is then classified into phonemes (a phoneme is defined as the subjective image employed by the brain of the listener in the process of speech recognition (p. 114); thus it is not strictly equivalent to the linguistic phoneme). Information distributed over an open syllable is employed in this classification process. At the next level, the string of phonemes is segmented, taking stress into account. Then the segmented string is interpreted as a sequence of words.

The Motor Theory of Speech Perception

Although motor theories of speech perception have been advanced by quite a number of theorists, the most explicit and reasoned statement

of the motor theory has been formulated by workers at Haskins Laboratories, namely F. S. Cooper, A. M. Liberman, D. P. Shankweiler, and others. For example, in an early discussion of some of their results (Cooper et al., 1952), the Haskins group advanced the motor theory.

The research as Haskins began with a search for invariants in speech--"A one-to-one correspondence between something half-hidden in the spectrogram and the successive phonemes of the message." (Cooper et al., 1952, p. 604). However, no acoustic invariant could be found for the individual phonemes. In fact, Cooper suggests that the perceived similarities and differences between speech sounds may correspond more closely to the similarities and differences in articulati than to the acoustic signal. As evidence for the simpler relation of perception and articulation, Cooper cites the complex relationship of the frequency of the burst of a stop consonant to the point of articulation: a burst of 1440 cps. is heard as /p/ before /i/ but as /k/ before /a/; conversely, bursts at different frequencies can be heard as the same consonant.

In connection with further work with synthetic speech, the Haskins group advanced the notion of categorial perception: perception of phonemes is different from perception of non-speech stimuli in that listeners can discriminate very little better than they can identify absolutely. An acoustic continuum is categorized into phonemes by listeners but a comparable non-speech continuum is not. Furthermore, listeners show discrimination peaks at phoneme boundaries when the stimulus is speech, but no such peaks in discrimination appear when the stimulus is a comparable non-speech continuum (Liberman, Harris,

Kinney, and Lane, 1957). These results, which are typically most clear-cut for stop consonants, are readily explained by the motor theory. It is argued that the gesture used in speech production is essentially invariant for the phoneme; therefore, perception is also invariant and categorial.

In their most detailed explication of the motor theory (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967), the Haskins group recapitulates the many arguments advanced for the motor theory and also specifies at what "level" production is made use of in perception. In their earlier work, the assumption was made that the production invariants were "motor commands" which were identical for each production of a given phoneme. In their latest statement, the idea of motor commands is retained and the theory is extended to higher-level neural signals which stand in a one-to-one relationship with other segments of the language:

> "In phoneme perception...the invariant is found far down in the neuromotor system, at the level of the commands to the muscles. Perception by morphophonemic, morphemic, and syntactic rules of the language would engage the encoding process at higher levels." (p. 454)

In this form, the motor theory becomes equivalent to analysis-by-synthesis, a theory of speech perception dependent on the use of rules in just such a way.

Analysis by Synthesis

Essentially, analysis by synthesis is a model of perception that depends on matching the incoming stimulus to an internally-generated pattern. When the internal pattern matches the stimulus, perception has been successful. As a model for speech perception, analysis by

synthesis has been extensively developed by Morris Halle and Kenneth N. Stevens.

An early version of the model (Halle and Stevens, 1964) is diagrammed in Fig. 4.

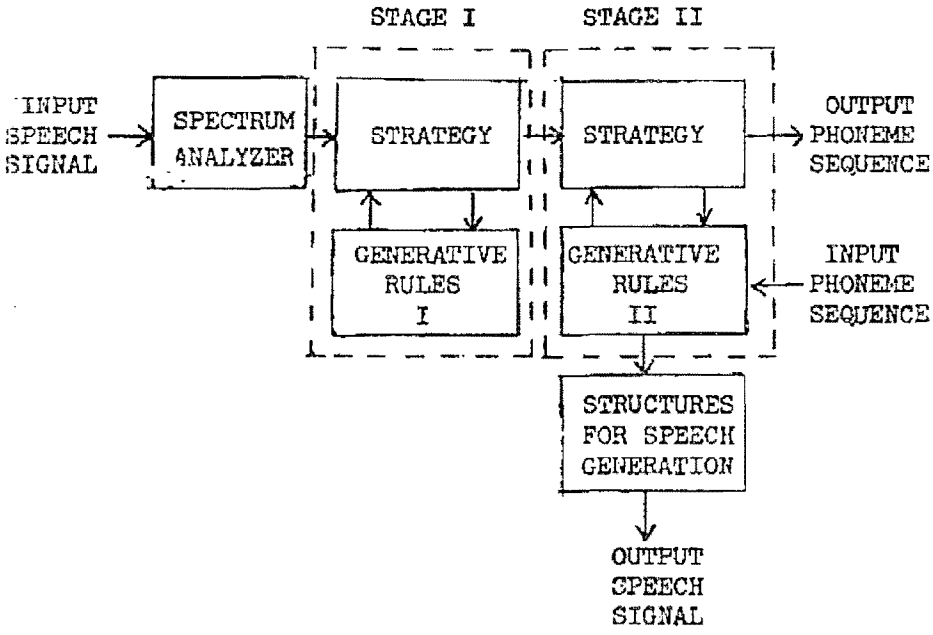


Fig. 4. Analysis by Synthesis model.
(Morris Halle and Kenneth N. Stevens, "Speech Recognition: a Model and a Program for Research," in The Structure of Language, ed. by Jerry A. Fodor and Jerrold G. Katz, 1964, by permission of Prentice-Hall).

The model depends on two analysis-by-synthesis loops. After a spectrum analysis, which in large part is a result of cochlear action, the first analysis-by-synthesis loop reduces the spectral representation of the acoustic input to a set of phonetic parameters. This is accomplished by matching the incoming spectrum to a spectrum produced by an internal synthesizer which has the ability to compute spectra when given phonetic parameters. In the second analysis-by-synthesis loop, the phonetic parameters are transformed to a sequence of phonemes. The second loop uses the generative rules that must also be employed

in speech production--rules that transform phonemes to phonetic parameters.

In a more recent statement of analysis-by-synthesis (Stevens and Halle, 1965), the analysis-by-synthesis model is integrated with linguistic concepts. The model is represented in Fig. 5.

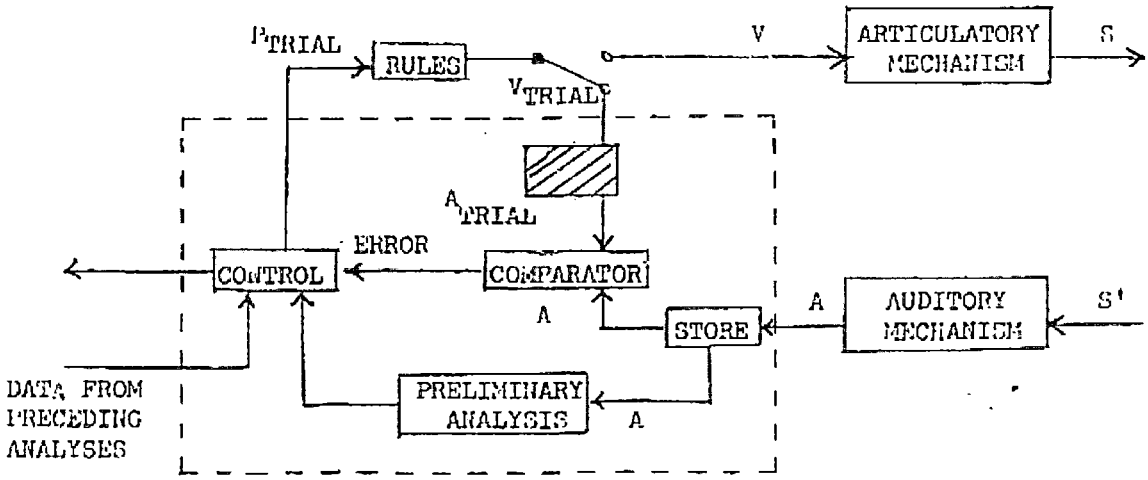


Fig. 5. Model for the speech-generating and speech-perception process. The dashed line encloses components of a hypothetical analysis-by-synthesis scheme for speech perception. (K. N. Stevens and M. Halle, "Remarks on Analysis by Synthesis and Distinctive Features," in Models for the Perception of Speech and Visual Form, 1965, by permission of M.I.T. Press.)

This model also claims that the mechanism employed in speech production is the same as the mechanism used in speech perception. Furthermore, the model employs abstract representations of words, coded in terms of distinctive features, and phonological rules, apparently identical to the rules found in the phonological component of a generative grammar.

The model operates in the following fashion. The auditory pattern derived from the acoustic input undergoes preliminary analysis; the exact nature of preliminary analysis is not specified in this model. On the basis of the preliminary analysis and contextual information, a

hypothesis is made concerning the abstract representation of the utterance. The proposed abstract representation is converted to an equivalent auditory pattern and compared with the pattern under analysis. If there is agreement, then the hypothesized abstract representation is judged to be correct, and processing at more abstract levels can proceed.

The function of the rules is to convert abstract representations to instructions to the vocal tract or to the equivalent auditory representation. Thus, these rules are more abstract than the motor commands postulated for the motor theory of speech perception.

Perceptual Strategies

The theory of perceptual strategies has been developed in close relation to transformational grammar. Perceptual strategies are techniques used by listeners to arrive at a segmentation of a sentence into deep structure units and to assign the proper grammatical function to each component. The theory is the result of research by M. Garrett, J. A. Fodor, and Thomas Bever. At the present, it is in a much more fluid state than the other theories discussed so far, so it seems appropriate to discuss the development of the theory, as well as its current status.

The early statements of the theory (Fodor and Bever, 1965; Garrett, Bever, and Fodor, 1966) were based on the phenomenon of click localization: when presented with a sentence with a superimpose click, the subject locates the click toward the nearest constituent boundary. Furthermore, subjects localize clicks correctly primarily when they occur on a constituent boundary. This phenomenon is

interpreted to mean that surface structure constituents form perceptual units, tending to resist interruption by extraneous material.

In later work, more detailed analysis of perceptual strategies followed. Fodor, Garrett, and Bever (Fodor and Garrett, 1967; Fodor, Garrett, and Bever, 1968) suggest that information about the properties of specific lexical items is employed by listeners. The listener selects the verb of the sentence and classifies it according to the possible deep structure configurations it can occur with; then the listener checks all these possible deep structure configurations to see if the surface structure he is presented with is a possible transformational version of the deep structure. In this process of selecting possible deep structures, the subject takes advantage of surface structure markers; for example, "to" implies that the verb must be able to take a "for...to" complementizer.

Later work also indicated that surface structure constituents were not directly related to perception (Bever, Lackner, and Kirk, 1969). Rather, the units of perception seem to be deep structure units.

The current status of the theory of perceptual strategies, as well as a summary of relevant research, has been presented by Bever (1970). In this article, Bever rejects the theory of derivational complexity. This theory claims that the perceptual complexity of a sentence is directly related to the number of transformations involved in its derivation. (A theory of analysis-by-synthesis at a syntactic level would imply derivational complexity.) But Bever finds that, in many cases, transformations are not related to perceptual complexity. First, transformational rules that delete structure do not add complexity; second, certain reordering transformations may even

simplify perception. For example, (1) is no more complex--and may even be simpler--than (2);

(1) It amazed Bill that John left early.

(2) That John left early amazed Bill.

Bever then proceeds to discuss several perceptual strategies employed by listeners. Some of these are the following.

a. When faced with a sentence, the listener isolates those adjacent phrases of surface structure which could correspond to a sentence in deep structure. The listener accomplishes this by segmenting together items that could be related as "actor, action, object...modifier."

b. Unless there is information to the contrary, the first noun...verb clause is treated as the main clause.

c. Constructions are related internally according to semantic constraints. Essentially, the listener selects the most likely semantic organization.

d. Any Noun-Verb-Noun sequence that is potentially a unit corresponds to "actor, action, object."

e. The special properties of function words and verbs are employed.

There is no need to give a complete list of proposed perceptual strategies, since all of them are proposed more or less tentatively. The general thrust of the theory, however, is this: to integrate perceptual strategies that are discovered to be applicable in language with other perceptual and cognitive processes, and to determine how language is related to other human cognitive abilities.

CHAPTER TWO

THE PERCEPTION OF SUB-PHONEMIC PHONETIC DIFFERENCES

In the models of speech perception discussed in the preceding chapter, it has been implicitly assumed that phonetic differences that are less than phonemic can have no linguistic significance, and that such differences can not be of any use to the listener. ("Phonemic" is to be understood here as "reliably signaling a difference in meaning.") This assumption follows directly from the traditional notion of a phoneme as a functional unit, distinct from all other such units. This view is also implicit in the notion of "categorial perception of phonemes" recently advanced by workers at Haskins Laboratories (Studdert-Kennedy, Liberman, Harris, and Cooper, 1970). On the other hand, phoneticians can develop an ability to notice small phonetic differences. And even ordinary listeners are sensitive to non-linguistic information that may be carried by sub-phonemic differences; for example, in identifying a particular speaker, sub-phonemic information is employed. However, speaker identification judgments are not linguistic and may be based on a great deal more information than on the fine phonetic details of an utterance.

In order to establish a "baseline" for perceptual units, it would be helpful to determine exactly how much use a subject can make of non-phonemic phonetic differences for linguistic judgments

A preliminary study related to this question was conducted by D. B. Fry (1968). Fry found that he was able to identify productions of the two words lax and lacks with no contextual information provided. The experiment was conducted in the following way: Fry prepared a tape by splicing copies of one production of lax and one production of lacks in random order. He then listened to the tape, and, after hearing each word, he pushed a button to identify it. Fry obtained both identificati scores and reaction time to the two words. He found, to his surprise, that he could identify the utterances correctly 96 times out of 100 (a statistically significant result). Furthermore, he found that the reaction time to lacks was faster than to lax, although the difference was not statistically significant.

Fry's study is quite tentative, so it is not proper to draw a generalization from it. Fry tested only one subject, himself, and only one supposedly-homophonous word pair. There are a number of possible explanations of the results that do not imply that listeners are generally aware of sub-phonemic differences. First, Fry is a very fine phonetician; therefore, he may be sensitive to distinctions which completely escape the ordinary listener. Second, he may have, by chance tested very distinctive productions of the two words; ordinarily, the two words may not be nearly so distinctive. Finally, it may be that some error in one or the other of the two words made them distinctive but not in a linguistic sense--there may have been some extraneous noise on the original recording of the utterance.

However, Fry's finding, if it reflects a general listener ability, has considerable implications for theories of speech perception. Therefore, it seemed desirable to replicate Fry's experiment with contro

over the variables mentioned above.

Method

Stimuli: Ten pairs of words were selected, each pair consisting of one monomorphemic and one bi-morphemic word of the same phonemic shape. Each pair of words composed a sub-list; within the sub-list, the two words were recorded in random order, each word appearing ten times. Each sub-list was introduced by two sentences in which the two words to be tested appeared in context. The following word pairs were tested: wade/weighed, hose/hoes, bard/barred, pact/packed, lax/lacks, baste/based, adds/adze, mist/missed, laps/lapse, and guest/guessed. The speaker was a male graduate student, a speaker of General American, whose home is in Connecticut.

The following procedure was employed to record the stimulus tape: for each production of each word to be recorded, the speaker was presented with a sketch picturing an activity suggestive of the word; underneath the sketch was a sentence employing the word, and descriptive of the sketch. The speaker was certain that under these circumstances he could produce the "correct word."

Two stimulus tapes were recorded; the second tape was a counter-balanced version of the first tape. On both tapes, words within lists were separated by five seconds; sub-lists were separated by ten seconds. Both tapes were recorded in a sound-proof recording booth, on an Ampex 350 tape recorder, at 7 1/2 i.p.s.

Subjects: Two groups of subjects participated in the experiment: 17 undergraduate students with no training in phonetics, and 12 graduate students in an introductory or advanced phonetics class.

The subjects were informed that the purpose of the experiment was to determine how quickly and how accurately people could identify words that sound very much the same. The subjects were instructed to respond as quickly as possible and to guess if they did not know which word they heard.

Procedure: The instrumentation is described in the accompanying diagra (Fig. 6).

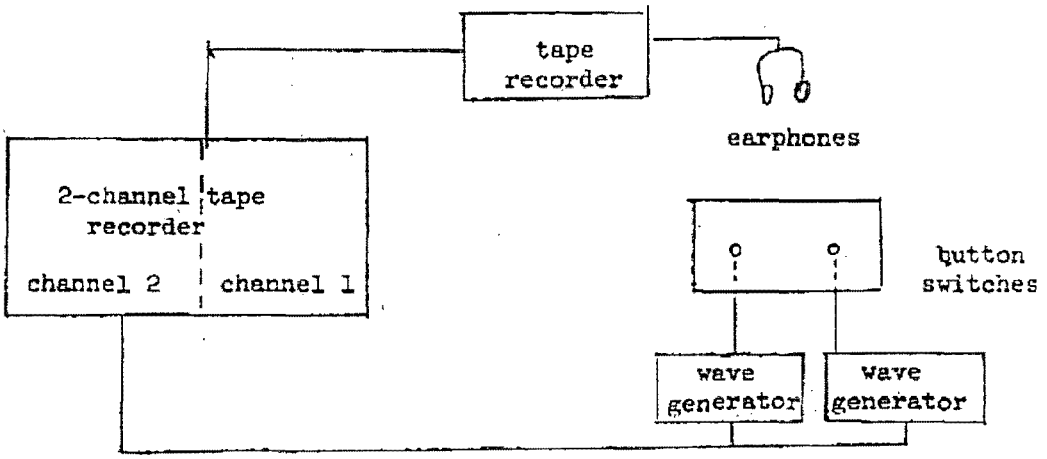


Fig. 6. Instrumentation for experiment testing the perception of sub-phonemic phonetic differences.

Each subject listened to the stimulus tape over earphones; he respond to each word by pushing one of two buttons, which were labeled, to identify which word he heard. The buttons were connected to two sig generators, one generating a sine wave, the other a square wave. Bo the stimulus tape and the subject's response were recorded on a two-channel tape recorder (Ampex 354) at 7 1/2 i.p.s. Thus both the rea time and the response were available for later analysis. Each subje responded to one complete list of 200 utterances. After the test, e subject was asked which pairs of words he felt he did well on and which pairs he felt he could not tell apart.

The tapes of each subject's performance were analyzed by computer. First, the voltages on each tape were digitized on a Radiation Inc. Analog Data Conversion System 152. The Ohio State University Instruction and Research Computer Center's IBM S/360 Mod 75 computer was used for further processing. The computer was programmed to determine changes in voltage. The transition from silence to voltage on the response channel was interpreted as the beginning of a response. The response was then categorized as either a sine wave or a square wave. The second channel containing voice was scanned to determine the transition from silence to voltage. This was construed as the beginning of a signal. The difference between the beginning of the signal and the beginning of the response was considered to be reaction time.[1]

---

[1]Measuring reaction time to speech stimuli, which exist in time, presents a problem not encountered with measuring reaction time to visual stimuli, namely at what point the subject can be said to begin to respond. The subject may begin to respond during the presentation of the word or after he has heard the entire word. On the other hand, reaction time can be measured either from the beginning or the end of the word. For this experiment, I have chosen to measure reaction time from the beginning of the word, in full awareness that either decision creates difficulties.

However, because of technical difficulties with the recordings, not all responses by every subject could be recovered.

Results

Identification: The over-all scores, given in Table 1, indicate that subjects do not seem to be able to identify the words correctly at significantly above chance levels. These results are repeated

TABLE 1
PER CENT CORRECT IDENTIFICATIONS

Word Pair	Total		List A	List B	Phonetics Students	Phonetically Untrained Students
	per cent correct	range of scores in per cent				
1. wade/ weighed	50.4	20-70	46.3	55.2	48.9	51.3
2. hose/ hoes	51.1	30-73	46.3	56.8	52.8	50.1
3. bard/ barred	50.6	30-75	53.6	47.1	52.1	49.6
4. pact/ packed	50.4	33-67	52.2	48.6	54.2	48.1
5. lax/ lacks	45.1	20-70	47.2	42.8	42.6	47.1
6. baste/ based	49.6	25-70	46.4	53.5	43.2	53.7
7. adds/ adze	46.2	25-75	43.5	49.0	46.5	45.3
8. mist/ missed	45.5	25-65	48.9	42.5	47.8	43.9
9. laps/ lapse	55.4	30-75	55.2	55.6	51.8	58.8
10. guest/ guessed	49.0	20-85	48.7	49.1	46.8	50.6

graphically in Fig. 7. Furthermore, phonetics students do not seem to perform significantly differently from phonetically untrained subjects.

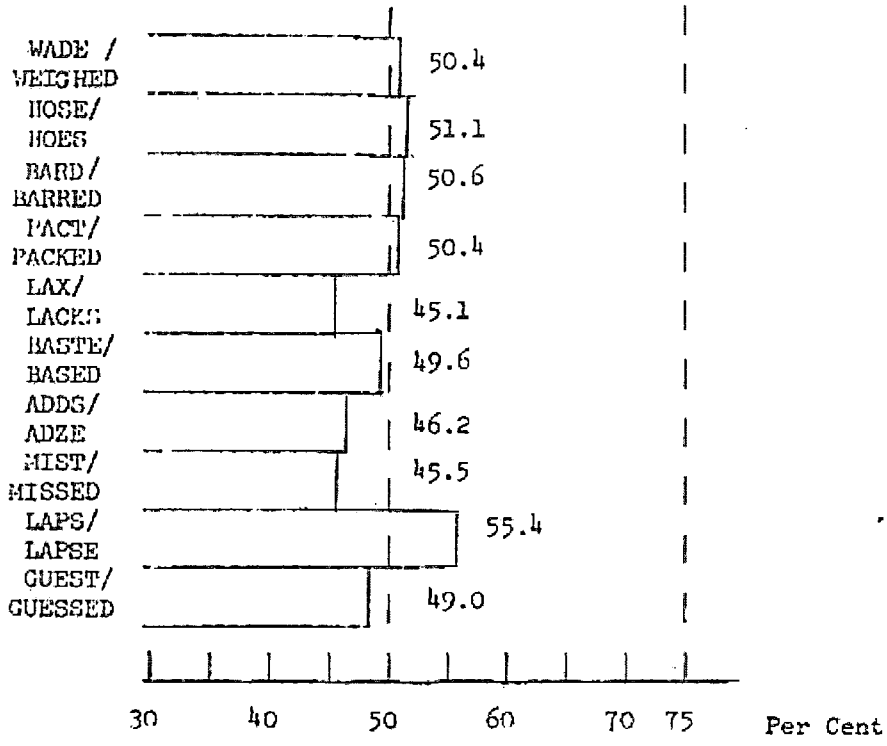


Fig. 7. Per Cent correct identifications for each word pair.

When the responses of the subjects to each production are analyzed, however, it appears that subjects are very consistent in their responses to some of the test items. Clearly consistent judgments (significant at .02 level or higher) for at least one production were obtained for the following pairs tested: weighed/wade, barred/bard, lax/lacks, baste/based, and mist/missed. Two pairs tested did not produce any significant agreement among subjects: hose/hoes and lapse/laps. Three pairs may or may not be considered significant; in each of these pairs, agreement in responses was reached for four productions at a .05 level of significance.

TABLE 2

CONSISTENCY OF SUBJECTS' RESPONSES
PER CENT OF S AGREEING IN RESPONSE B
(underlined scores are significant at .02 level)

List A

production number	wade	hose	bard	pact	lax	baste	adds	mist	lapse	guest
1	61.5	16.7	66.7	54.5	46.2	85.7	53.3	100.0	54.5	25.0
2	53.8	69.2	69.2	36.4	33.3	33.3	50.0	30.0	81.8	41.7
3	41.7	66.7	23.1	81.8	23.1	42.9	58.3	50.0	54.5	58.3
4	53.8	61.5	76.9	36.4	50.0	42.9	33.3	40.0	81.8	66.7
5	50.0	76.9	58.3	81.8	33.3	50.0	41.7	50.0	36.4	45.5
6	53.8	50.0	58.3	54.5	25.0	50.0	41.7	60.0	36.4	45.5
7	50.0	53.8	15.4	70.0	46.2	57.1	50.0	66.7	45.5	58.3
8	46.1	69.2	75.0	54.5	61.5	66.7	50.0	70.0	72.7	41.7
9	45.5	46.1	38.5	81.8	53.8	71.4	16.7	60.0	63.6	66.7
10	38.5	46.1	61.5	81.8	61.5	66.7	75.0	20.0	45.5	50.0
11	69.2	76.9	30.8	30.0	18.2	18.2	33.3	30.0	45.5	41.7
12	58.3	66.7	30.8	50.0	53.8	72.7	75.0	70.0	50.0	58.3
13	38.46	46.1	61.5	27.3	53.8	50.0	33.3	66.7	54.5	16.7
14	30.8	46.1	84.6	63.6	41.7	61.5	54.5	50.0	45.5	50.0
15	63.6	46.1	46.1	45.5	30.8	78.5	58.3	50.0	27.3	58.3
16	69.2	58.3	46.1	54.5	50.0	58.3	25.0	80.0	54.5	58.3
17	69.2	53.8	61.5	63.6	46.2	28.6	83.3	55.5	45.5	66.7
18	38.5	61.5	46.1	45.5	46.2	57.1	58.3	40.0	54.5	58.3
19	46.1	61.5	15.4	45.5	58.3	78.5	58.3	40.0	63.6	50.0
20	61.5	61.5	38.5	63.6	38.5	33.3	41.7	70.0	54.5	50.0

List B

production number	wade	hose	bard	pact	lax	baste	adds	mist	lapse	guest
1	63.6	45.5	27.3	60.0	58.3	16.7	75.0	63.6	71.4	42.9
2	54.5	55.5	45.5	20.0	50.0	36.4	50.0	45.5	28.6	69.2
3	63.6	55.6	54.5	55.6	50.0	58.3	25.0	63.6	50.0	61.5
4	54.5	50.5	18.2	70.0	54.5	33.3	25.0	63.6	50.0	50.0
5	60.0	27.3	66.7	50.0	45.6	75.0	75.0	54.5	71.4	57.1
6	54.5	60.0	60.0	55.6	63.6	50.0	16.7	36.4	57.1	21.4
7	36.4	36.4	30.0	50.0	41.7	66.7	50.0	50.0	42.9	64.3
8	54.5	81.8	10.0	44.4	45.6	50.0	41.7	50.0	61.5	69.2
9	54.5	10.0	50.0	77.8	83.3	20.0	83.3	45.5	42.9	28.6
10	100.0	45.5	45.5	60.0	41.7	54.5	50.0	72.7	64.3	35.7
11	66.7	55.5	63.6	60.0	54.5	25.0	58.3	70.0	50.0	28.6
12	50.0	70.0	36.4	60.0	36.4	50.0	36.3	27.3	50.0	21.4
13	36.4	45.5	63.6	25.0	66.7	41.7	54.5	27.3	30.8	42.9
14	63.6	36.4	30.0	50.0	33.3	75.0	58.3	81.8	46.2	21.4
15	60.0	55.6	18.2	30.0	33.3	45.5	50.0	27.3	71.4	61.5
16	63.6	55.6	54.5	50.0	41.7	33.3	54.5	27.3	28.6	53.8
17	54.5	63.6	30.0	57.1	66.7	66.7	45.5	45.5	64.3	64.3
18	18.2	40.0	36.4	70.0	22.2	25.0	50.0	45.5	57.1	42.9
19	72.7	27.3	40.0	55.6	63.6	72.7	63.6	63.6	42.9	57.1
20	72.7	60.0	45.5	20.0	66.7	50.0	25.0	63.6	35.7	58.3

The consistency of subjects' responses is represented in Table 2.

Even when subjects are highly consistent in agreeing on a particular response, they do not necessarily identify the word correctly; the identification scores for utterances for which subjects agree on one response (at .02 level) are still at chance level (57% correct).

Subject Interview: The mean identification score for the word pair judged easiest and for the most difficult word pair was calculated. The score represents each subject's performance in relation to his judgment of ease and difficulty, and thus does not represent performance on any one word pair. The differences found were not statistically significant, but did lie in an interesting direction: both phonetically trained and phonetically untrained subjects performed better on the word pairs they considered easy than on the word pairs they considered difficult.

TABLE 3

SUBJECTS' PERFORMANCE IN RELATION TO JUDGMENTS OF EASE AND DIFFICULTY

	Word Pair Judged Easiest (% Correct)	Word Pair Judged Most Difficult (% Correct)
All Subjects	53.10	46.01
Phonetics Students	51.60	49.20
Phonetically Untrained Students	54.10	43.80

Furthermore, subjects show a fair amount of agreement in judging which pairs of words are difficult and which are easy. Table 4 shows

the number of times each word pair was judged easy and the number of times each word pair was judged difficult.

TABLE 4

EASE AND DIFFICULTY OF WORD PAIRS AS JUDGED BY SUBJECTS

Word pair	Number of times judged easy	Number of times judged difficult
wade/weighed	6	4
hose/hoes	3	7
bard/barred	5	1
pact/packed	1	2
lax/lacks	1	3
baste/based	3	2
adds/adze	1	5
mist/missed	3	0
laps/lapse	3	3
guest/guessed	3	1

Reaction time: Reaction time was not determined for all subjects. As Tables 5 to 8 show, reaction time was quite slow for all subjects and to all word pairs. There is no significant systematic difference in reaction time between correct and incorrect responses.

Reaction time to productions labeled consistently is quite variable. When the reaction time to consistently labeled productions is compared with the mean reaction time for that word pair, the differences in reaction time are in no way systematic. When the differences are statistically significant, however, then reaction time is longer to the consistently labeled production. These data are presented in Table 9.

When reaction time to mono-morphemic and to bi-morphemic words is examined, there is some tendency for reaction time to be shorter

TABLE 5

REACTION TIME, IN SECONDS, FOR SUBJECTS WITH TRAINING IN PHONETICS, FOR THE PAIRS WADE/WEIGHED, HOSE/HOES, BARD/BARRED, PACT/PACKED, AND LAX/LACKS

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	wade / weighed		hose / hoes		bard / barred		pact / packed		lax / lacks	
D.G.										
all responses	1.180	.118	1.103	.250	.970	.152	1.229	.207	1.265	.222
correct responses	1.178	.129	1.170	.167	.916	.083	1.223	.257	1.289	.203
incorrect responses	1.183	.113	1.058	.292	1.024	.188	1.234	.169	1.245	.244
L.S.										
all responses	1.098	.199	1.181	.234	1.003	.180	--	--	1.192	.303
correct responses	.994	.153	1.187	.227	1.042	.168	--	--	1.147	.225
incorrect responses	1.159	.203	1.176	.252	.930	.194	--	--	1.220	.355
Z.B.										
all responses	--	--	--	--	--	--	--	--	--	--
correct responses	--	--	--	--	--	--	--	--	--	--
incorrect responses	--	--	--	--	--	--	--	--	--	--
S.Z.										
all responses	1.261	.398	1.814	.623	1.234	.546	1.651	.505	--	--
correct responses	1.348	.377	1.806	.649	1.320	.627	1.597	.518	--	--
incorrect responses	1.213	.419	1.822	.633	1.075	.348	1.698	.523	--	--

TABLE 8

REACTION TIME, IN SECONDS, FOR SUBJECTS WITH TRAINING IN PHONETICS, FOR THE PAIRS BASTE/BASED, ADDS/ADZE, MIST/MISSED, LAPS/LAPSE, AND GUEST/GUESSED

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	baste / based		adds / adze		mist / missed		laps / lapse		guest / guessed	
D.G.										
all responses	1.096	.189	.957	.149	1.115	.227	1.137	.198	1.019	.081
correct responses	1.009	.063	.970	.199	1.158	.227	1.101	.158	.930	--
incorrect responses	1.143	.219	.948	.115	1.071	.178	1.190	.248	1.033	.078
L.S.										
all responses	--	--	--	--	.983	.159	1.369	.358	1.186	.294
correct responses	--	--	--	--	.973	.138	1.353	.417	1.188	.291
incorrect responses	--	--	--	--	.993	.189	1.402	.235	1.185	.324
Z.B.										
all responses	1.542	.440	1.468	.563	1.620	.525	1.589	.516	--	--
correct responses	1.465	.454	1.491	.586	1.603	.515	1.571	.406	--	--
incorrect responses	1.598	.443	1.442	.581	1.644	.599	1.597	.587	--	--
S.Z.										
all responses	1.263	.506	1.081	.330	--	--	1.243	.464	1.265	.509
correct responses	1.267	.292	1.049	.230	--	--	1.246	.413	1.088	.257
incorrect responses	1.261	.619	1.124	.445	--	--	1.241	.529	1.353	.590

TABLE 7

REACTION TIME, IN SECONDS, FOR PHONETICALLY UNTRAINED SUBJECTS FOR THE PAIRS WADE/WEIGHED, HOSE/HOES, BARD/BARRED, PACT/PACKED, AND LAX/LACKS

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	wade / weighed		hose / hoes		bard / barred		pact / packed		lax / lacks	
1 all resp.	1.513	.281	1.377	.192	1.263	.241	1.376	.186	1.194	.327
correct resp.	1.403	.164	1.310	.176	1.214	.197	1.397	.199	1.180	.299
incorrect resp.	1.560	.312	1.421	.197	1.336	.295	1.350	.177	1.211	.375
2 all resp.	1.967	.516	2.458	.945	1.901	.642	2.382	.780	1.751	.618
correct resp.	1.725	.516	2.125	.502	2.181	.964	2.340	.718	1.631	.551
incorrect resp.	2.304	1.190	2.863	1.212	1.780	.439	2.431	.892	1.976	.714
3 all resp.	1.663	.816	2.150	.941	1.635	.515	1.687	.486	1.380	.569
correct resp.	1.443	.744	2.158	.941	1.570	.423	1.719	.521	2.045	.926
incorrect resp.	2.320	.734	--	--	1.690	.317	1.638	.473	1.158	.215
4 all resp.	2.056	.651	1.549	.667	1.323	.599	--	--	1.042	.639
correct resp.	2.061	.623	1.769	.768	1.227	.565	--	--	1.176	.793
incorrect resp.	2.050	.739	1.329	.502	1.498	.673	--	--	.922	.483
5 all resp.	1.734	.552	--	--	1.723	.460	2.139	.507	--	--
correct resp.	1.778	.532	--	--	1.741	.537	2.105	.459	--	--
incorrect resp.	1.640	.646	--	--	1.709	.425	2.162	.579	--	--
6 all resp.	--	--	1.867	.525	1.635	.268	--	--	--	--
correct resp.	--	--	1.778	.458	1.585	.191	--	--	--	--
incorrect resp.	--	--	1.993	.622	1.662	.309	--	--	--	--
7 all resp.	1.376	.228	1.778	.476	--	--	1.699	.272	1.628	.348
correct resp.	1.363	.168	1.779	.487	--	--	1.739	.290	1.443	.250
incorrect resp.	1.385	.278	1.776	.210	--	--	1.636	.258	1.814	.352
8 all resp.	1.972	.326	1.917	.336	--	--	2.364	.813	2.627	.664
correct resp.	1.910	.318	1.934	.329	--	--	2.778	.921	2.242	.305
incorrect resp.	2.057	.338	1.891	.376	--	--	2.123	.669	2.742	.709

TABLE 8

REACTION TIME, IN SECONDS, FOR PHONETICALLY UNTRAINED SUBJECTS, FOR THE PAIRS BASTE/BASED, ADDS/ADZE, MIST/MISSED, LAPS/LAPSE, AND GUEST/GUESSED

(mean and standard deviation; significantly different means for correct vs. incorrect responses are underlined)

Subject	baste / based		adds / adze		mist / missed		laps / lapse		guest / guessed	
1 all resp.	1.489	.267	--	--	--	--	--	--	--	--
correct resp.	1.431	.268	--	--	--	--	--	--	--	--
incorrect resp.	1.527	.271	--	--	--	--	--	--	--	--
2 all resp.	2.232	.543	--	--	--	--	--	--	--	--
correct resp.	2.305	.473	--	--	--	--	--	--	--	--
incorrect resp.	2.123	.652	--	--	--	--	--	--	--	--
3 all resp.	1.828	.859	1.280	.384	1.782	.729	1.725	.593	1.970	.476
correct resp.	2.163	.887	1.325	.247	1.793	.684	1.815	.604	1.984	.341
incorrect resp.	1.358	.617	1.267	.430	1.771	.828	1.235	.120	1.963	.548
4 all resp.	1.697	.754	1.523	.637	--	--	--	--	--	--
correct resp.	1.855	.735	1.400	.466	--	--	--	--	--	--
incorrect resp.	1.348	.749	1.625	.780	--	--	--	--	--	--
5 all resp.	--	--	--	--	--	--	--	--	1.845	.718
correct resp.	--	--	--	--	--	--	--	--	1.826	.566
incorrect resp.	--	--	--	--	--	--	--	--	1.857	.836
6 all resp.	--	--	.960	.185	.927	.134	1.103	.397	1.070	.296
correct resp.	--	--	.966	.154	.874	.149	.996	.264	1.136	.332
incorrect resp.	--	--	.954	.220	.951	.127	1.296	.547	.978	.241
7 all resp.	1.937	.385	1.742	.249	1.848	.373	1.652	.408	--	--
correct resp.	1.971	.441	1.748	.268	1.908	.434	1.506	.151	--	--
incorrect resp.	1.862	.246	1.729	.230	1.779	.305	1.972	.618	--	--
8 all resp.	2.073	.541	--	--	--	--	2.134	.311	2.135	.515
correct resp.	1.901	.385	--	--	--	--	2.283	.310	2.049	.468
incorrect resp.	2.345	.665	--	--	--	--	1.909	.129	2.292	.604

TABLE 9

REACTION TIME, IN SECONDS, TO PRODUCTIONS LABELED CONSISTENTLY
(reaction times significantly different from mean are underlined)

Subject	Mean RT for word pair	RT to production	Mean RT for word pair	RT to production			Mean RT to word pair	RT to production	Mean RT to word pair	RT to production	Mean RT to word pair	RT to production
		10 wade		7 bard	14 bard	19 barred		11 lax		1 based		1 mist
3	1.663	.720										
4	2.056	1.700										
5	1.734	1.681										
6	--	--										
7	1.376	1.527										
8	1.972	2.100										
D.G.			.970	.870	.970	.950	1.265	1.300	1.096	1.600	1.115	1.730
L.S.			1.003	1.194	.944	.994	--	--	--	--	.983	1.090
Z.B.			--	--	--	--	--	--	1.542	2.281	1.620	--
S.Z.			1.234	1.760	--	1.080	--	--	1.263	3.070	--	--
1			1.263	1.420	1.250	1.250	1.194	.950	1.489	1.160	--	--
2			1.901	1.070	1.950	3.650	1.751	3.370	2.232	2.870	--	--

to the bi-morphemic word, as Fry discovered. The differences, however, are not statistically significant. These data are presented in Tables 10 to 12.

Acoustic analysis: In order to discover the acoustic cues that subjects were employing to arrive at consistent labeling, spectrograms were made of all productions that were labeled consistently. Spectrograms were also made of some productions for each word pair that were labeled at random, and of the production that immediately preceded the consistently labeled production. Spectrograms were made on a Kay Electric Company Sonagraph.

It was found that subjects were employing two types of cues: slight differences in consonant quality and differences in vowel duration. For the word pairs baste/based, mist/missed, and lax/lacks, subjects were responding to a slight difference in the fricative [s]. The consistently labeled productions had more energy, at all frequencies, in the fricative than the productions that were labeled at random.

The word pairs wade/weighed and bard/barred were labeled consistently on the basis of vowel duration. However, subjects apparently were not responding to absolute differences in vowel duration, but to the duration of a vowel compared to the duration of the vowel of the preceding production. Thus a production [beɹd] would be labeled barred if it followed a production with a perceptibly shorter vowel; it would be labeled bard if it followed a production with a perceptibly longer vowel. It did not matter whether the word was intended as "bard" or "barred."

TABLE 10

REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS WEIGHED/WADE, HOES/HOSE, BARRED/BARD, PACKED/PACT

(mean and standard deviation; significantly different means are underlined)

Subj.	weighed		wade		hoes		hose		barred		bard		packed		pact	
DG	1.129	.091	1.227	.152	1.083	.110	1.258	.181	.908	.080	.924	.095	1.206	.262	1.245	.289
LS	1.106	.094	.910	.138	1.152	.127	1.210	.285	1.039	.215	1.046	.115	--	--	--	--
ZB	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SZ	1.255	.384	1.535	.403	1.765	.530	1.817	.718	1.152	.448	1.522	.799	1.830	.521	1.286	.383
1	1.320	.064	1.445	.191	1.377	.115	1.270	.205	1.179	.186	1.264	.223	1.305	.100	1.508	.241
2	1.652	.479	1.878	.626	2.278	.594	1.942	.333	2.990	.436	1.778	.778	2.473	.634	2.182	.854
3	1.464	.493	1.403	1.207	1.870	.398	2.590	1.612	1.608	.478	1.420	--	2.108	.604	1.498	.507
4	1.940	.747	2.142	.586	2.090	.439	1.576	.902	1.542	.718	.965	.211	--	--	--	--
5	1.798	.693	1.775	.317	1.745	.426	1.826	.567	1.860	.475	1.025	--	1.742	.233	2.468	.222
6	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
7	1.441	.113	1.312	.199	1.908	.574	1.606	.372	1.466	.010	1.764	.092	1.916	.315	1.632	.245
8	1.873	--	1.970	.329	2.074	.441	1.864	.278	--	--	--	--	2.299	.886	3.415	.546

TABLE 11
REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS LACKS/LAX, BASED/BASTE, ADDS/ADZE
(mean and standard deviation; significantly different means are underlined)

Subject	lacks		lax		based		baste		adds		adze	
DG	1.256	.269	1.330	.096	1.003	.056	1.017	.084	1.018	.134	.923	.261
LS	1.289	.069	.935	.205	--	--	--	--	--	--	--	--
ZB	--	--	--	--	1.679	.411	1.467	.408	1.459	.700	1.524	.556
SZ	--	--	--	--	1.255	.092	1.272	.354	1.104	.307	.980	.067
1	1.446	.213	.958	.011	1.555	.262	1.308	.239	--	--	--	--
2	1.453	.429	1.896	.676	2.190	.498	2.420	.461	--	--	--	--
3	2.045	.926	--	--	2.647	1.215	1.800	.422	1.500	--	1.150	--
4	1.042	.725	1.400	1.015	2.292	.525	1.332	.669	1.327	.418	1.510	.693
5	--	--	--	--	--	--	--	--	--	--	--	--
6	--	--	--	--	--	--	--	--	.862	.101	1.070	.126
7	1.435	.320	1.305	--	1.948	.449	1.997	.482	1.673	.148	1.879	.398
8	1.966	--	2.380	.269	2.175	.295	1.672	.296	--	--	--	--

TABLE 12

REACTION TIME, IN SECONDS, TO THE MONO-MORPHEMIC AND BI-MORPHEMIC WORDS MISSED/MIST, LAPS/LAPSE, GUESSED/GUEST

(mean and standard deviation; significantly different means are underlined)

Subject	missed		mist		laps		lapse		guessed		guest	
DG	1.070	.150	1.246	.271	1.088	.143	1.115	.184	.930	--	--	--
LS	.927	.189	1.010	.102	1.110	.130	1.515	.473	1.188	.291	--	--
ZB	1.461	.515	1.660	.563	1.197	--	1.758	.346	--	--	--	--
SZ	--	--	--	--	.966	.168	1.665	.205	1.030	.226	1.127	.318
1	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--
3	1.735	.007	1.816	.836	1.630	.241	1.968	.787	2.155	.191	1.870	.407
4	--	--	--	--	--	--	--	--	--	--	--	--
5	--	--	--	--	--	--	--	--	--	--	1.885	.677
6	1.010	.127	.783	.075	.910	.101	1.038	.084	1.260	.495	1.043	.173
7	2.295	--	1.859	.438	1.463	.106	1.531	.174	--	--	--	--
8	--	--	--	--	2.171	.339	2.339	.311	1.908	.368	2.219	.559

Discussion

To a great extent, the results of this experiment are negative. Subjects can not identify the word pairs correctly. They do not perform better on the word pairs they consider easy than on the word pairs they consider difficult. And no inferences can be drawn from the reaction time except that, because the reaction time is very slow, the subjects find it difficult to decide which word they have heard.

However, subjects seem to be aware of at least some sub-phonemic information since they label some word pairs consistently, even though not correctly. Faced with the task of the experiment, subjects develop a strategy for making use of fine phonetic detail. In this manner they arrive at some consistent labelings. But since the identifications based on this strategy are equally likely to be correct or incorrect, the strategy can not be considered to be part of ordinary speech perception.

Thus the results of the experiment imply that even though subjects may become aware of sub-phonemic differences, they do not know what linguistic use to make of them.

CHAPTER THREE

THE PERCEPTION OF OBSTRUENT CLUSTERS

Studies dealing with the perception of order of non-speech sounds indicate that perceiving the order of sounds of short duration is quite problematic. Hirsch (1959) reported that, after considerable practice, subjects could perceive the order of two sounds correctly if the onset of the sounds was separated by 15 to 20 msec. For stimuli, Hirsch used tones and bursts of noise 500 msec. in duration as well as clicks. Hirsch concludes that the minimal temporal interval required for perception of order is independent of the duration of the sound (within the limits of the experiment) and of the quality of the sound.

Broadbent and Ladefoged (1959) found that, at first, subjects could not perceive the order of sounds unless the onset of the sounds was separated by 150 msec.; with considerable training, a 30 msec. separation became adequate for accurate perception of order. Broadbent and Ladefoged used three different stimuli: a "hiss," high frequency noise of 120 msec. duration; a "pip," an 800 cps sine wave of 30 msec. duration, and a "buzz," a 171 cps square wave of 30 msec. duration.

Both these experiments involved the perception of the order of only two elements. However, the task is much more difficult when the

subject has to determine the order of three or more elements. Several experiments involving the ordering of more than two sounds are reported by Warren and Warren (1970). In the first experiment, subjects were asked to determine the order of three sounds--a hiss, a tone, and a buzz, each lasting 200 msec.--which were repeated over and over without pauses. The subjects performed no better than chance. When the order of four sounds--a high tone, a low tone, a buzz, and a hiss, each lasting 200 msec.--was to be judged, the duration of each item had to be increased to between 300 and 700 msec. for half of the subjects to identify the sequence correctly. In the last experiment, the subjects were asked to judge the order of four 200 msec. vowel segments, cut from productions of extended vowels and spliced together without pauses. The subjects performed no better than chance. Identification of order became possible only when a 50 msec. silent interval was introduced between the vowels.

These experiments show that subjects have considerable difficulty in perceiving the order of sounds. However, listeners have no comparable difficulty with the order of elements in perceiving speech, even though many speech sounds are of quite short duration. Words like tax and task, ax and ask are normally perceived correctly, even though the duration of the consonants in the cluster is close to the minimum discovered in the Hirsch experiment. A reasonable estimate of the duration of p, t, and k is 51 msec., 30 msec. and 36 msec., respectively (Lehiste, 1970). These figures are derived from Estonian short voiceless stops.

It is, of course, a common observation that children have difficulty with such clusters; aks is a very common child pronunciation

of ask, for example. And historically, such clusters have been prone to metathesis.[1] Still, adults seem to have no trouble with these clusters in the ordinary use of speech.

[1]It may be that the sporadic occurrence of metathesis, found in historical change, could be better explained by examining errors in perception rather than errors in production, which has been the traditional starting point for discussing language change.

The observation that children have trouble with obstruent clusters but adults do not could imply that the adults' proficiency is a result of considerable practice. Both the Broadbent and Ladefoged, and Hirsch experiments show that the perception of order improves with practice. Analogously, the adults' proficiency could be a result of practice acquired in the course of language learning. However, it is also possible, and has been suggested by a number of theorists, that some special mechanisms are employed in the perception of consonant clusters. Thus Broadbent and Ladefoged report that the introspective feeling, developed in judging order, is that two items become differentiated on the basis of over-all quality rather than order. They suggest that the perceptual mechanism operates on discrete samples of perceptual information; when two items fall into the same sample their order has to be inferred on some other basis. On the basis of the Broadbent and Ladefoged and Hirsch experiments, Neisser (1967) argues that a listener gradually learns to distinguish a cluster like ts from a cluster like st, rather than perceiving a sequence of t followed by s, or s followed by t. He implies that such clusters are perceptual units to the listener, not normally analyzed further.

Wickelgren's idea of context sensitive coding, presented in detail in Chapter One (Wickelgren, 1969a, 1969b), can also explain the fact that adults easily perceive a sequence of consonants correctly. When a listener is presented with a consonant cluster, e.g. sk, he knows that it is composed of two elements , but he does not encode these elements in order; rather, the cluster is coded as an unordered sequence, with each element identified for what precedes and follows it. Schematically, the coding would be something like the following: $s^{k}{}_{\#}$ $\#^{s}{}_{k}$. These elements can be assembled in the correct order, and the listener can arrive at the intended sequence.

The perception of obstruent clusters is an interesting problem for empirical study, particularly since it is related to the almost universally accepted notion that the minimal unit in speech perception is the phoneme. Both Neisser's suggestion and Wickelgren's theory, if substantiated, would argue against this view.

An experiment was designed to investigate the perceptual mechanisms employed in the perception of obstruent clusters. By observing the pattern of confusions of obstruent clusters in the presence of noise, it is possible to make some inferences about the perceptual mechanisms underlying the perception of these clusters.

Method

Stimuli: Fifteen pairs of English words were selected which differed from each other only in the order of obstruents in a cluster. Five pairs of words ended in the obstruent cluster ps/sp; five ended in ts/st; five ended in ks/sk. For each obstruent cluster, there was one pair of two-syllable words; in addition, each obstruent cluster

appeared at least once with no morpheme boundary in the cluster. The full list of words is reproduced below:

apse	Blatz	ax
asp	blast	ask
lips	mats	tax
lisp	mast	task
Capsian	blitzer	axing
Caspian	blister	asking
claps	boots	Max
clasp	boost	mask
raps	coats	bricks
rasp	coast	brisk

Three lists were constructed. On each list, each word appeared two times in random order; the order was arrived at by using a table of random numbers. Thus each list consisted of 60 words; each consonant cluster appeared on each list ten times.

The speaker was a male, with a medium-pitch voice, from Akron, Ohio. Before recording, the speaker practiced for some time so that he could produce the stressed vowel of each word at a constant intensity. This was accomplished by monitoring the v.u. meter on the tape recorder. When the speaker was producing the words at a constant intensity, the actual recording was made, monitoring each production to keep the intensity at a constant level. The three lists were recorded in a sound-proof recording booth on an Ampex 350 tape recorder, at 7 1/2 i.p.s. Words were separated by 2.5 seconds; after every five words, there was a gap of 5 seconds.

The stimulus tape was made by re-recording the master tape while adding "white" noise produced by a Grayson-Stadler noise generator.

The instrumentation is shown in the accompanying diagram (Fig. 8).

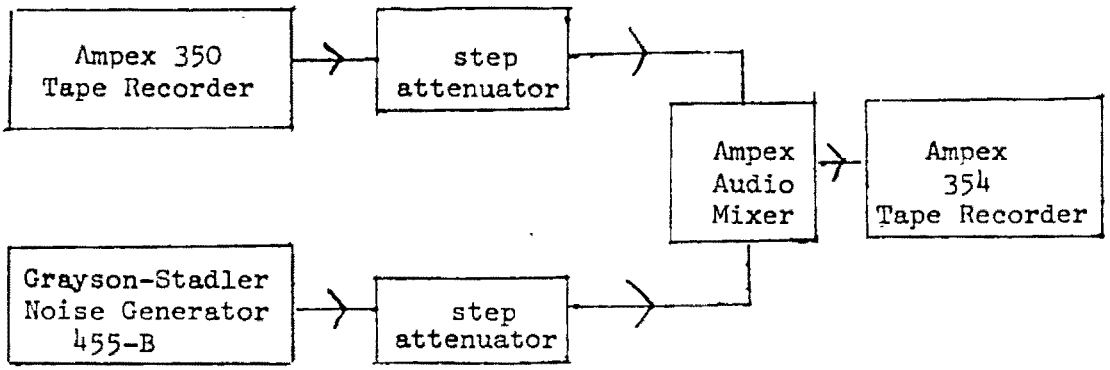


Fig. 8. Instrumentation for adding noise to stimulus tape.

Three different signal-to-noise ratios were employed for the three lists: the first list was re-recorded at a signal-to-noise ratio of 0 d.b.; the second list was re-recorded at a signal-to-noise ratio of +12 d.b.; the third list was recorded at a signal-to-noise ratio of -6 d.b.

Subjects: Nineteen subjects participated in the experiment. All were members of The Ohio State University linguistics department and native speakers of English.

Procedure: The experiment was conducted as a listening test. Before the test, subjects were instructed to write what they heard, and to guess if necessary; they were told to expect some unusual words, and these words were shown to them. For the test, the stimulus tape was played on a tape recorder while the subjects listened over earphones, and wrote what they heard on an answer sheet. Each subject listened to the entire tape (3 lists), and thus responded to 180 stimulus words.

In addition, five subjects took the test a second time. In the second test, the listening conditions were identical to those of the

first test, but the subjects were instructed to say what they heard. The subjects' spoken response and the stimulus tape were recorded on separate channels of an Ampex 354 tape recorder.

The subjects' responses were tabulated in the form of confusion matrices. The answers were scored only for the perception of the obstruent clusters. Thus, if the stimulus word was raps, but the subject wrote laps, he was scored correct.

The response tapes of the five subjects who gave spoken responses were processed by an Elema-Schönander Mingograf, each channel of the tape being represented as an oscillogram on a separate channel of the Mingograf. The paper speed was 100 mm/sec.

Reaction time was determined by measuring from the onset of the stimulus word to the onset of the response, and from the end of the stimulus word to the onset of the response. There was no difficulty in measurement when the signal-to-noise ratio was +12 d.b. When the signal-to-noise ratio was 0 d.b., measurements from the stimulus word had to be made from the vowel rather than from the consonants. Reaction time could not be determined when the signal-to-noise ratio was -6 d.b.

Results

Confusions: The results are presented in the accompanying confusion matrices (Tables 13 to 51). Each cell of the matrices shows the number of times the stimulus consonant cluster, given at the beginning of the row, was identified as the consonant cluster given in the column heading. Correct responses lie on the diagonal. In addition,

the percent of all the responses of each row that lie in a particular cell is given for each cell. A.I. (articulation index) gives the ratio of correct identifications for each matrix.

Tables 13 to 15 give confusion matrices for all responses. As is to be expected, the higher the noise is, in relation to the signal, the more confusion errors occur. It can be observed that, for all consonant clusters, the most common error is a reversal of the consonant cluster. Furthermore, the stop-fricative cluster is perceived correctly more often than the corresponding fricative-stop cluster. This effect may result from the higher frequency of stop-fricative clusters in English.

The pattern of confusions for written responses (Tables 16 to 18) and for spoken responses (Tables 19 to 21) is essentially the same. Thus, there is no advantage to spoken responses, and spoken responses do not produce a different pattern of confusions.

Tables 22 to 27 present the confusion matrices for two-syllable words. The articulation index is slightly higher for two-syllable words, but the confusion patterns remain essentially the same. There is some tendency to confuse p and k clusters only with each other, and not with t clusters; however, this is probably due to other differences in the two-syllable words tested, i.e., a different vowel and a different final consonant.

Tables 28 to 45 present confusion matrices for all test words with a given vowel. The most common confusion, for all vowels, is still a reversal of the consonant cluster. There is only one exception to this tendency; when the vowel is [ɪ], p clusters tend to be confused with t clusters about as much as with each other.

TABLE 13
ALL RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8599

	TS	ST	PS	SP	KS	SK
TS	81.9 181	9 20	2.3 5		1.4 3	5.4 12
ST	6.5 15	78.8 182		1.3 3	.9 2	12.5 29
PS	1.7 4	.8 2	86.2 206	8.4 20	2.5 6	.4 1
SP	.4 1	3.5 8	3.5 8	77.3 177		15.3 35
KS		.4 1	1.3 3	.9 2	95.2 219	2.2 5
SK	1.3 3		1.3 3	.8 2	.4 1	96.2 226

TABLE 14
ALL RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .4896

	TS	ST	PS	SP	KS	SK
TS	54.3 109	33.3 67	2.9 6	.5 1	.9 1	7.9 16
ST	45.3 87	39.6 76	5.2 10	.5 1	.5 1	8.9 17
PS	10.2 19	7 13	44.6 83	27.4 51	3.3 6	7.5 14
SP	14.5 27	7 13	23.7 44	37.1 69	4.3 8	13.4 25
KS	9.1 16	3.4 6	1.7 3	2.8 5	67.6 119	15.4 27
SK	7.7 13	2.9 5	14.3 24	5.4 9	17.9 30	51.8 87

TABLE 15

ALL RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .3874

	TS	ST	PS	SP	KS	SK
TS	57.2 88	29.2 45	1.9 3	1.3 2	5.8 9	4.6 7
ST	55.7 78	32.2 45	1.4 2	1.4 2	5 7	4.3 6
PS	10.4 18	4.6 8	36.4 63	32.5 46	9.2 16	6.9 12
SP	19.7 29	13.6 20	13.6 20	29.9 44	11.6 17	11.6 17
KS	8.5 12	3.5 5	9.2 13	7.1 10	38.4 54	33.3 47
SK	4.9 5	2.9 3	14.7 15	15.7 16	24.5 25	37.3 38

TABLE 16
ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8529

	TG	ST	PS	SP	KS	SK
TG	81.4 139	10.5 18	2.3 4		1.2 2	4.6 8
ST	7.7 14	77.9 141		1.7 3	1.1 2	11.6 21
PS	1.6 3	1.1 2	85 160	9.6 18	2.1 4	.6 1
SP	.6 1	2.8 5	3.9 7	76.9 137		15.8 28
KS		.6 1	1.6 3	1.1 2	94.5 169	2.2 4
SK	1.7 3		1.7 3	.5 1	.5 1	95.6 176

TABLE 17
ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .5006

	TG	ST	PS	SP	KS	SK
TG	55.5 87	32.5 51	3.1 5	.6 1	1.3 2	7 11
ST	44.5 69	41.1 60	5.5 8	.7 1	.7 1	7.5 11
PS	7.9 11	7.9 17	42.5 59	29.5 41	2.9 4	9.3 13
SP	11.4 16	9.2 13	24.8 35	36.9 52	4.9 7	12.8 18
KS	8.9 12	3.7 5	2.2 3	1.5 2	66.6 92	14.9 20
SK	5.6 7	3.1 4	11.1 14	5.6 7	17.5 22	57.1 72

TABLE 18
ALL WRITTEN RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .4121

	TS	ST	PS	SP	KS	SK
TS	56.7 63	36 34	1.8 2	1.8 2	7.2 8	1.8 2
ST	51 52	40.2 41	.98 1	1.96 2	2.94 3	2.94 3
PS	8.86 11	5.64 7	37.9 47	31.4 39	9.7 12	6.5 8
SP	15.3 16	13.3 14	16.2 17	34.3 36	12.4 13	8.6 9
KS	8.2 8	4.1 4	11.2 11	7.2 7	37.8 37	31.6 31
SK	5.8 4	4.4 3	14.5 10	15.9 11	20.3 14	39.2 27

TABLE 19
TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: +12 d.b.

AI: .8849

	TG	ST	PS	SP	KS	SK
TS	84 42	4 2	2 1		2 1	8 4
ST	2 1	82 41				16 8
PS	1.9 1		90.4 46	3.9 2	3.9 2	
SP		5.9 3	1.9 1	78.5 40		13.7 7
KS					98.2 50	1.9 1
SK				1.9 1		98.2 50

TABLE 20
TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: 0 d.b.

AI: .4548

	TG	ST	PS	SP	KS	SK
TS	50 22	36.4 16	2.3 1			11.3 5
ST	47.9 22	34.8 16	4.35 2			13.02 6
PS	17 8	4.25 2	51 24	21.3 10	4.25 2	2.13 1
SP	24.4 11		20 9	37.8 17	2.2 1	15.6 7
KS	9.6 4	2.4 1		6.7 3	64.4 27	16.7 7
SK	14.3 6	2.4 1	23.8 10	4.8 2	19 8	35.7 15

TABLE 21
TOTAL SPOKEN RESPONSES--SIGNAL TO NOISE RATIO: -6 d.b.

AI: .3266

	TS	ST	PS	SP	KS	SK
TS	58.2 25	25.6 11	2.3 1		2.3 1	11.6 5
ST	68.5 26	10.5 4	2.6 1		10.5 4	7.9 3
PS	14.3 7	20.4 1	32.7 16	34.7 17	8.2 4	8.2 4
SP	31 13	14.3 6	6.7 3	19 8	9.6 4	19 8
KS	9.3 4	2.3 1	4.7 2	6.9 3	39.6 17	37.2 16
SK	3 1		15.2 5	15.2 5	33.3 11	33.3 11

TABLE 22
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .9598

	TS	ST	PS	SP	KS	SK
TS	88.9 32	8.3 3			2.8 1	
ST	2.7 1	97.3 36				
PS			97.5 39	2.5 1		
SP		2.6 1		97.4 37		
KS			2.9 1		97.1 34	
SK				2.6 1		97.4 37

TABLE 23
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .5730

	TS	ST	PS	SP	KS	SK
TS	40.7 11	51.9 14				7.4 2
ST	30.6 11	66.7 24	2.7 1			
PS			37.9 11	58.7 17		3.4 1
SP			25.8 8	67.7 21		6.5 2
KS	4 1			8 2	64 16	24 6
SK	3.3 1		6.7 2	16.7 5	10 3	63.3 19

TABLE 24
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .95

	TS	ST	PS	SP	KS	SK
TS	100 10					
ST		100 10				
PS			100 10			
SP			10 1	90 9		
KS					90 9	10 1
SK				10 1		90 9

TABLE 25
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .636

	TS	ST	PS	SP	KS	SK
TS	50 5	50 5				
ST	50 5	50 5				
PS			70 7	30 3		
SP			10 1	70 7	20 2	
KS				16.7 1	83.3 5	
SK			11.1 1	11.1 1	11.1 1	66.7 6

TABLE 26
WRITTEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, Capsian/Caspian, asking/axing)
AI: .4524

	TS	ST	PS	SP	KS	SK
TS	50 7	50 7				
ST	50 10	45 9	5 1			
PS			40.6 13	50 16	3.1 1	6.3 2
SP			19.1 4	71.4 15		9.5 2
KS	4.8 1		28.5 6	23.8 5	23.8 5	19.1 4
SK				44.4 8	11.1 2	44.4 8

TABLE 27
SPOKEN RESPONSES FOR TWO-SYLLABLE WORDS
SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, Capsian/Caspian, axing/asking)
AI: .3953

	TS	ST	PS	SP	KS	SK
TS	80 4	20 1				
ST	100 5					
PS			44.4 4	44.4 4		11.1 1
SP	10 1	20 2	10 1	50 5		10 1
KS				20 1	20 1	60 3
SK			11.1 1	33.3 3	22.2 2	33.3 3

TABLE 28
SPOKEN RESPONSES FOR [I]--SIGNAL TO NOISE RATIO: +12 d.b.
(blister/blitzer, lips/lisp, brisk/bricks)
AI: .9000

	TS	ST	PS	SP	KS	SK
TS	100 10					
ST		100 10				
PS	10 1		70 7		20 2	
SP		20 2		70 7		10 1
KS					100 10	
SK						100 10

TABLE 29
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: +12 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .8683

	TS	ST	PS	SP	KS	SK
TS	70 14	5 1	5 1			20 4
ST	4.8 1	57.1 12				38.1 8
PS			95.1 39	4.9 2		
SP		2.4 1	2.4 1	80.6 33		14.6 6
KS					97.6 40	2.4 1
SK				2.4 1		97.6 40

TABLE 30
SPOKEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: +12 d.b.
(coats/coast, boots/boost)
AI: .9487

	TS	ST	PS	SP	KS	SK
TS	90 18	5 1			5 1	
ST		100 19				

TABLE 31
SPOKEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: 0 d.b.
(blister/blitzer, lips/lisp, bricks/brisk)
AI: .4655

	TS	ST	PS	SP	KS	SK
TS	50 5	50 5				
ST	50 5	50 5				
PS	30 3	10 1	20 2	20 2	20 2	
SP	40 4		20 2	40 4		
KS					90 9	10 1
SK			25 2		50 4	25 2

TABLE 32
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: 0 d.b.
(matz/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .4412

	TS	ST	PS	SP	KS	SK
TS	28.6 4	28.6 4	7.1 1			35.7 5
ST	27.8 5	27.8 5	11.1 2			33.3 6
PS	13.5 5	2.7 1	59.5 22	21.6 8		2.7 1
SP	20 7		20 7	37.1 13	2.9 1	20 7
KS	12.5 4	3.1 1		9.4 3	56.3 18	18.7 6
SK	17.6 6	2.8 1	23.6 8	5.4 2	11.8 4	38.2 13

TABLE 33
SPOKEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: 0 d.b.
(coast/coats, boost/boats)
AI: .5000

	TS	ST	PS	SP	KS	SK
TS	65 13	35 7				
ST	66.7 12	33.3 6				

TABLE 34
SPOKEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: -6 d.b.
(blister/blitzer, lips/lisp, bricks/brisk)
AI: .2791

	TS	ST	PS	SP	KS	SK
TS	80 4	20 1				
ST	83.3 5				16.7 1	
PS	40 4	10 1	10 1	10 1	20 2	10 1
SP	66.7 6	11 1	11.1 1	11.1 1		
KS		11.1 1			44.4 4	44.4 4
SK			33.3 1		33.3 1	66.7 2

TABLE 35
SPOKEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: -6 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .3136

	TS	ST	PS	SP	KS	SK
TS	45 9	20 4	5 1		5 1	25 5
ST	50 7		7.2 1		21.4 3	21.4 3
PS	7.7 3		38.5 15	41 16	5.1 2	7.7 3
SP	21.2 7	15.2 5	6.1 2	21.2 7	12.1 4	24.2 8
KS	11.8 4		5.9 2	8.8 3	38.2 13	35.3 12
SK	3.5 1		13.8 4	17.2 5	34.5 10	31 9

TABLE 36
SPOKEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: -6 d.b.
(coats/coast, boots/boost)
AI: .4444

	TS	ST	PS	SP	KS	SK
TS	66.7 12	33.3 6				
ST	77.8 14	22.2 4				

TABLE 37
WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: -6 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .4117

	TS	ST	PS	SP	KS	SK
TS	55.3 21	13.2 5	2.6 1	2.6 1	21 8	5.3 2
ST	34.6 9	30.9 8	3.8 1	7.7 2	11.5 3	11.5 3
PS	1.1 1	3.3 3	42.5 39	35.8 33	11.9 11	5.4 5
SP	6.7 5	9.3 7	18.7 14	40 30	16 12	9.3 7
KS	10.3 7	4.4 3	13.2 9	8.8 6	39.8 27	23.5 16
SK	3.4 2	5.2 3	15.5 9	19 11	19 11	37.9 22

TABLE 38
WRITTEN RESPONSES FOR [I]--SIGNAL TO NOISE RATIO: -6 d.b.
(lips/lisp, bricks/brisk, blister/blitzer)
AI: .3094

	TS	ST	PS	SP	KS	SK
TS	33.3 6	55.6 10	5.6 1	5.6 1		
ST	55.6 10	44.4 8				
PS	31.3 10	12.5 4	25 8	18.7 6	3.1 1	9.4 3
SP	36.6 11	23.3 7	10 3	20 6	3.3 1	6.7 2
KS	3.3 1	3.3 1	6.7 2	3.3 1	33.3 10	50 15
SK	18.2 2		9.1 1		27.3 3	45.4 5

TABLE 39
WRITTEN RESPONSES FOR [u] AND [oʊ]--SIGNAL TO NOISE RATIO: -6 d.b.
(boots/boost, coats/coast)
AI: .5398

	TS	ST	PS	SP	KS	SK
TS	65.5 36	44.5 19				
ST	56.9 33	53.1 25				

TABLE 40
WRITTEN RESPONSES FOR [ɪ]--SIGNAL TO NOISE RATIO: 0 d.b.
(lips/lisp, bricks/brisk, blister/ blitzer)
AI: .5515

	TS	ST	PS	SP	KS	SK
TS	50 16	43.8 14				6.2 2
ST	22.8 8	77.2 27				
PS	19.4 6	16.1 5	32.3 10	29 9		3.2 1
SP	17.8 5	17.8 5	21.5 6	42.9 12		
KS	5.3 1				94.7 18	
SK	5 1		5 1	5 1	45 9	40 8

TABLE 41
WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: 0 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/ rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .4991

	TS	ST	PS	SP	KS	SK
TS	59.3 45	18.4 14	6.6 5	1.3 1	2.6 2	11.8 9
ST	40.3 23	26.3 15	12.3 7	1.8 1		19.3 11
PS	4.6 5	5.6 6	45.4 49	29.7 32	3.6 4	11.1 12
SP	9.7 11	7.1 8	25.6 29	35.4 40	6.2 7	16 18
KS	9.6 11	4.4 5	2.6 3	1.7 2	64.3 74	17.4 20
SK	5.6 6	3.8 4	12.3 13	5.6 6	12.3 13	60.4 64

TABLE 42
WRITTEN RESPONSES FOR [ʊ] AND [oʊ]--SIGNAL TO NOISE RATIO: 0 d.b.
(boost/boots, coast/coats)
AI: .4271

	TS	ST	PS	SP	KS	SK
TS	53 26	47 23				
ST	63 34	33.3 18	1.9 1		1.9 1	

TABLE 43
WRITTEN RESPONSES FOR [æ]--SIGNAL TO NOISE RATIO: +12 d.b.
(mats/mast, Blatz/blast, ax/ask, apse/asp, Max/mask, tax/task, raps/rasp, claps/clasp, Capsian/Caspian, asking/axing)
AI: .8173

	TS	ST	PS	SP	KS	SK
TS	68.8 27	6.9 3	9.3 4		2.4 1	18.6 8
ST	17.9 12	43.3 29		4.5 3	2.9 2	31.4 21
PS	.6 1	1.3 2	86.5 133	11 17		.6 1
SP	.7 1	2.1 3	4.2 6	75.4 107		17.6 25
KS			2.1 3	1.4 2	93.7 134	2.8 4
SK	2.1 3		2.1 3	.7 1	.7 1	94.4 138

TABLE 44
WRITTEN RESPONSES FOR [I]--SIGNAL TO NOISE RATIO: +12 d.b.
(lips/lisp, bricks/brisk, blister/blitzer)
AI: .8909

	TS	ST	PS	SP	KS	SK
TS	78.6 33	19 8			2.4 1	
ST	2.9 1	97.1 33				
PS	5.8 2		79.4 27	2.9 1	11.9 4	
SP		5.6 2	2.8 1	83.3 30		8.3 3
KS		2.8 1			97.2 35	
SK						100 38

TABLE 45
WRITTEN RESPONSES FOR [u] AND [ou]--SIGNAL TO NOISE RATIO: +12 d.b.
(boost/boots, coast/coats)
AI: .9518

	TS	ST	PS	SP	KS	SK
TS	91.9 79	8.1 7				
ST	1.2 1	98.8 79				

For both p and t, the second formant transition would be negative before [ɪ] (as opposed to k). Perhaps this fact accounts for the confusion.

Tables 46 to 51 present confusions for bi-morphemic words. Apparently, the presence of a morpheme boundary does not deter confusions; rather, mono-morphemic and bi-morphemic words produce similar confusion patterns.

Reaction time: Reaction time was compared for the two different signal-to-noise conditions, for words ending in different consonant clusters, and for correct vs. incorrect responses.

Reaction time was significantly faster when the signal-to-noise ratio was +12 d.b., than when the signal-to-noise ratio was 0 d.b.

As can be seen in Table 52, reaction time was consistently faster for correct responses than for incorrect responses, although the difference did not always reach statistical significance.

When the reaction time to the individual consonant clusters is examined, the reaction time is significantly slower to words ending in ps, sp, and sk clusters when the signal-to-noise ratio is 0 d.b. When the signal-to-noise ratio is +12 d.b., reaction time is significantly slower only to words ending in ps clusters.[2] (Table 53).

---

[2]This difference may be a result of the frequency of the words. For example, apse is not even listed in An English Word Count (Wright, 1965).

Finally, the reaction time to two-syllable words, when measured from the beginning of the word, is about the same as the reaction time to one-syllable words. When measured from the end of the word, the

TABLE 46

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: +12 d.b.

(lips, claps, naps, bricks, coats, mats, boots)

AI: .8391

	TS	ST	PS	SP	KS	SK
TS	79.3 73	9.8 9	4.3 4			6.6 6
PS	1.9 2		83.7 87	10.6 11	3.8 4	
KS		2.9 1			97.1 33	

TABLE 47

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: 0 d.b.

(lips, claps, naps, bricks, coats, mats, boots)

AI: .4798

	TS	ST	PS	SP	KS	SK
TS	45 36	35 28	5 4		1.3 1	13.7 11
PS	13.5 10	14.9 11	39.2 29	28.4 21	1.4 1	2.6 2
KS	5.3 1				94.7 18	

TABLE 48

WRITTEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: -6 d.b.

(lips, claps, naps, bricks, coats, mats, boots)

AI: .4702

	TS	ST	PS	SP	KS	SK
TS	63.9 46	25 18	2.8 2			8.3 6
PS	20.9 17	6.2 5	34.6 28	29.7 24	4.9 4	3.7 3
KS		3.1 1	6.2 2	3.1 1	40.7 13	46.9 15

TABLE 49
SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: +12 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .8116

	TS	ST	PS	SP	KS	SK
TS	72.5 21	6.9 2	3.4 1		3.4 1	13.8 4
PS	3.3 1		83.3 25	6.7 2	6.7 2	
KS					100 10	

TABLE 50
SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: 0 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .4769

	TS	ST	PS	SP	KS	SK
TS	48.4 14	31 9			3.4 1	17.2 5
PS	26.9 7	7.7 2	30.8 8	26.9 7	7.7 2	
KS					90 9	10 1

TABLE 51
SPOKEN RESPONSES FOR BI-MORPHEMIC WORDS--SIGNAL TO NOISE RATIO: -6 d.b.
(lips, claps, raps, bricks, coats, mats, boots)
AI: .4194

	TS	ST	PS	SP	KS	SK
TS	50 13	26.9 7	3.8 1		3.8 1	15.5 4
PS	26.9 7		30.8 8	26.9 7	11.6 3	3.8 1
KS	10 1				50 5	40 4

TABLE 52
REACTION TIME, IN MSEC., FOR CORRECT AND INCORRECT RESPONSES

Consonant Cluster	Signal to Noise Ratio: 0 d.b.				Signal to Noise Ratio: +12 d.b.			
	Correct		Incorrect		Correct		Incorrect	
	Beginning	End	Beginning	End	Beginning	End	Beginning	End
1 syllable words:								
TS	1011	744	989	779	887	646	1012	739
ST	1209	939	957	734	863	619	946	664
2 syllable words:								
TS	954	519	1198	827	862	377	-	-
ST	1169	809	1102	746	853	376	-	-
1 syllable words:								
PS	1038	845	1167	980	960	747	1247	1063
SP	1042	839	1211	986	918	673	809	582
2 syllable words:								
PS	984	474	1025	514	875	343	-	-
SP	1095	633	919	412	876	312	930	360
1 syllable words:								
KS	1067	866	1074	862	837	619	-	-
SK	931	692	1194	945	918	679	-	-
2 syllable words:								1190
KS	1156	690	1018	655	826	319	1610	695
SK	1008	489	1420	923	808	287	1175	

TABLE 53
REACTION TIME, IN MSEC., TO CONSONANT CLUSTER
(Mono-syllabic words only)

	Signal to Noise Ratio: 0 d.b.		Signal to Noise Ratio: +12 d.b.	
	Beginning	End	Beginning	End
TS	998	773	913	667
ST	1035	771	904	629
PS	1109	914	981	772
SP	1092	878	891	649
KS	1037	835	837	620
SK	1138	890	918	679

reaction time is much shorter to two-syllable words. Apparently, subjects begin to respond to the two-syllable words before they hear the whole word, probably as soon as they hear the medial consonant cluster.

Discussion

The finding that has the most bearing on the perception of consonant clusters is that reversal errors are the most common errors. This finding is counter to the idea that the phoneme is the minimal perceptual unit; if consonant clusters are perceived "phoneme-by-phoneme," then, when a listener hears the consonant cluster sp, he first hears s and then he hears p. Given that he hears these in a particular order, there is no reason for him to reverse that order. Granted, he might on occasion forget the order, but there is no reason to suppose that he would be more likely to forget the order of the consonants than to forget one of the consonants; thus, reversal errors would be no more common than substitution errors. However, that is clearly not the case: reversal errors are much more common. This finding implies that some special perceptual mechanisms must be postulated for the perception of consonant clusters.

Broadbent and Ladefoged's suggestion appears of doubtful validity, not because the consonant cluster data contradict it, but for other reasons. As has already been pointed out by Neisser, a listener is not limited to an invariant time-determined chunk of input that he can process. This is implied by the ability of listeners to perceive correctly speech that is speeded up. Broadbent and Ladefoged would have to claim that order errors would become more common, and involve more segments, as speech is speeded up, since each "time chunk"

would contain more segments. But that this is not the case seems clear from personal experience with record players.

Neisser's suggestion, that a consonant cluster is a perceptual unit, and Wickelgren's suggestion that a consonant cluster is coded in terms of some element very much like an allophone, are both compatible with the data.

If consonant clusters are perceptual units, then clearly a ps cluster is most similar to a sp cluster. If this is so, then, when the signal is degraded by the addition of noise, the items that are most similar to each other will be confused most; thus, reversal errors will be most likely.

If a consonant cluster is coded in terms of allophones, then the allophone of s before p will be slightly different, acoustically, from the allophone of s after p. This difference, however, will be the most subtle part of the signal; particularly, it will be smaller than the acoustic information differentiating consonants from each other. These small acoustic differences will be the first to disappear when the signal is degraded by noise; consequently, reversal errors will be the most common in a degraded signal.

Thus, either Neisser's or Wickelgren's suggestion will account for the observed result.

CHAPTER FOUR

SYNTACTIC UNITS IN PERCEPTION

Experiments involving the localization of "clicks" in sentences have been used by Bever, Fodor, and others (Fodor and Bever, 1965; Bever, Lackner and Kirk, 1969) to examine syntactic units in perception. The experiments are based on a phenomenon discovered by Ladefoged and Broadbent (1960) that subjects have great difficulty localizing a click in speech, when the click and speech are presented simultaneously.

At first, the "click" experiments seemed to support the view that syntactic constituents were perceptual units: when asked to locate a click, subjects tended to move it towards a constituent boundary. A theory of perception was developed to explain the phenomenon: a subject could pay attention to one thing at a time, he could either process speech or the click; subjects would not interrupt perceptual units of speech; consequently, subjects would tend to locate the click between perceptual units.

However, the click-locating task, as defined in the early experiments involved a complex interaction of perception and memory, since the subject had to remember the sentence he had just heard, remember where the click had occurred, and locate the click in a written version of the sentence.

Reaction time is a response measure that is more directly linked

to perception in that the subject is not required to remember the click location. But when reaction time to clicks was measured, it was found that reaction time was not shortest to clicks located in constituent boundaries, as the theory would predict, and furthermore, reaction time did not seem to be related to the syntactic structure of a sentence (Abrams and Bever, 1970).

In order to explain this development, Abrams and Bever suggest a different model of attention in speech perception; they argue that the latency of the response to the click is a function of a subject's over-all attention to sensory input. At the beginning of a clause, the subject must pay attention to the input very closely, hence his reaction time to clicks is fast. At the end of clauses, the subject can already predict much of what is to come, so he does not have to pay much attention, and his reaction time to clicks is slower.

But it is also possible that constituent structure is not directly involved in perception, but is a result of perceptual analysis. It is possible that reaction time is a function of the suprasegmental structure of a sentence, as suggested by Dr. Lehiste (personal communication).

An experiment was designed to test a part of this hypothesis, namely to determine whether reaction time to clicks is affected by their relation to stressed elements.

Method

Stimuli: Ten sentences were selected to serve as stimuli. Each sentence was recorded two times in random order. Sentences were separated by a pause of 5 seconds. The recording was made in a

sound-proof booth and an Ampex 350 tape recorder, at 7 1/2 i.p.s.

The speaker was male, with a medium pitched voice. He was instructed to say the sentences clearly and naturally.

One click was placed in each sentence. There were four types of click location: in a stressed vowel, in an unstressed vowel, in the consonant preceding a stressed vowel, and in the consonant preceding an unstressed vowel. In addition, one click was located in a constituent boundary. The clicks were produced by a capacitor discharge, triggered by the release of a key. The click so produced was a single spike, with a very rapid rise and decay. The duration of each click was approximately 25 msec.

The stimulus tape was made by re-recording the sentences on one channel of an Ampex 354 tape recorder and recording the click, at the appropriate time, on the second channel. In addition, five clicks were recorded on the stimulus tape before the clicks which were associated with sentences, to determine each subject's reaction time to non-speech stimuli.

The sentences employed, and the location of the clicks, are given below. For convenience, the location of clicks in both productions of the sentence is shown in one written version of the sentences. The complex sentences are taken from the study conducted by Abrams and Bever (1970); the simple sentences are taken from a study conducted by Lehiste (1971).

1. That the matter was dealt with fast, was a surprise to Harry.
2. Since she was free that day, her friends asked her to come

3. My sleep was disturbed.
4. By making his plan known, Jim brought out the objections of everybody.
5. Speed kills.
6. Any student who is bright but young, would not have seen it.
7. The speed was controlled.
8. Sleep refreshes.
9. If you did call up Bill, I thank you for your trouble.
10. After the dry summer of that year, some of the crops were completely lost.

Click location was verified by inspecting the oscillograms, produced by two channels of an Elema-Schönander Mingograf, representing the two channels of the stimulus tape.

Subjects: Eleven subjects participated in the experiment. All were members of the Ohio State University linguistics department.

Procedure: Each subject listened to the stimulus tape two times. The first time, he was instructed to listen to the sentences and to push a key as quickly as he could when he heard the click. The key triggered a capacitor discharge which was recorded directly on one channel of an Elema-Schönander Mingograf. Simultaneously, the channel of the stimulus tape which contained the clicks was recorded on another channel of the Mingograf. The instrumentation is shown in the accompanying diagram (Fig. 9). Paper speed was 100 mm per second.

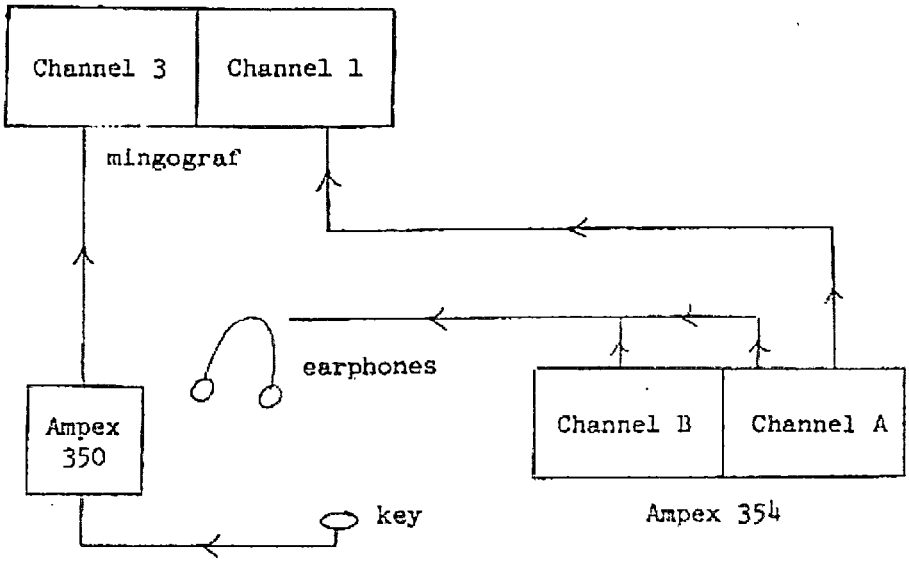


Fig. 9. Instrumentation for "click" experiment.

Immediately after the first test, the subject listened to the tape again. This time, he was provided with a written copy of each sentence and asked to mark the location of each click.

Reaction time to clicks was determined by measuring from the peak of the stimulus click to the onset of the response.

Results

The reaction time to clicks was compared for four conditions: when the click occurred in a stressed vowel, when it occurred in an unstressed vowel, when it occurred in a consonant preceding a stressed vowel, and when it occurred in a consonant preceding an unstressed vowel. The results are presented in Table 54 and in Fig. 10 to 12. Fig. 10 shows the reaction time to a click embedded in a consonant preceding a stressed vowel,and in a consonant preceding an unstressed vowel. For all but one subject, the reaction time is faster to the click preceding an unstressed vowel. Fig. 11 shows reaction time to

TABLE 54
MEAN REACTION TIME TO CLICKS (IN MSEC.)

Subject	Click Location					Non-speech Click
	Stressed Vowel	Consonant Preceding Stressed Vowel	Unstressed Vowel	Consonant Preceding Unstressed Vowel	Constituent Boundary	
1	333	336	293	278	230	230
2	276	242	226	195	190	200
3	235	244	237	218	390	180
4	255	241	198	150	170	170
5	233	239	249	240	240	200
6	524	557	582	505	600	390
7	236	181	144	145	120	110
8	406	383	316	313	460	280
9	241	283	206	200	210	150
10	236	230	224	220	250	235
11	163	165	175	175	165	140
For all subjects	285	281	259	240	275	

[illegible] in stressed vowels and to clicks embedded in [illegible] vowels. For six subjects, the reaction time is faster [illegible] an unstressed vowel; for the other subjects, the reaction [illegible]ally the same.

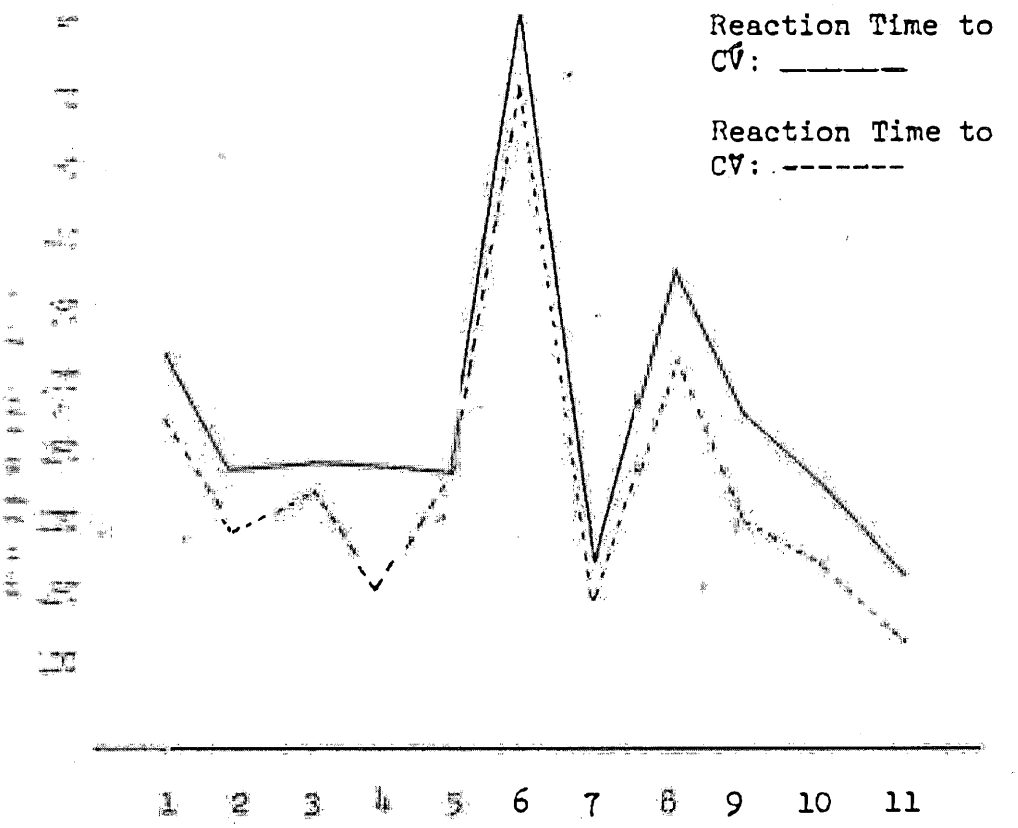


Fig. 10. Reaction time to clicks in consonants preceding stressed vowels and to clicks in consonants preceding unstressed vowels.

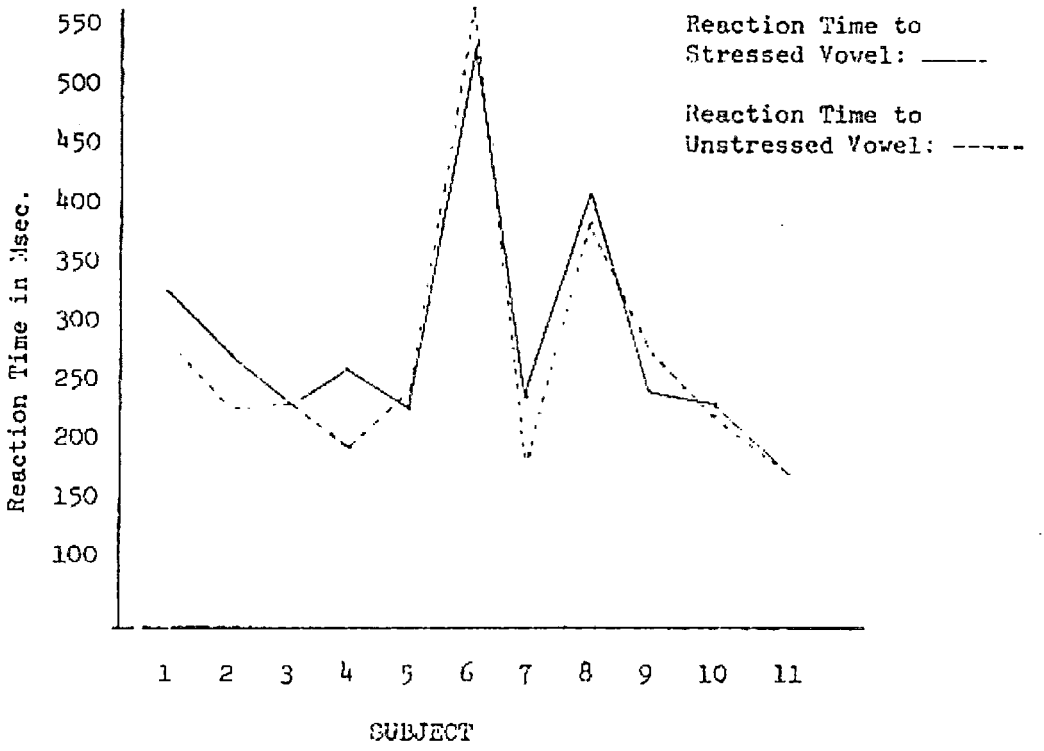


Fig. 11. Reaction time to clicks in stressed vowels and in unstressed vowels.

Although the differences are not always statistically significant, the tendency is clear: reaction time to clicks is affected by their location in relation to stressed elements. Reaction time to a click is longest when the click is in the vicinity of a stressed element, either in a stressed vowel or in a consonant preceding a stressed vowel. Reaction time is shorter when the click is in the vicinity of an unstressed element, either in an unstressed vowel or in a consonant preceding an unstressed vowel.

The reaction time to clicks located in constituent boundaries is quite variable. For some subjects, it is very short in this condition, approaching the reaction time to non-speech stimuli. For other subjects,

it is quite long, longer than the reaction time to clicks in any other condition.

Reaction time to non-speech clicks is short in all cases, implying that reacting to a click in a speech context is more complex than simply reacting to a click. These results are presented in Fig. 12.

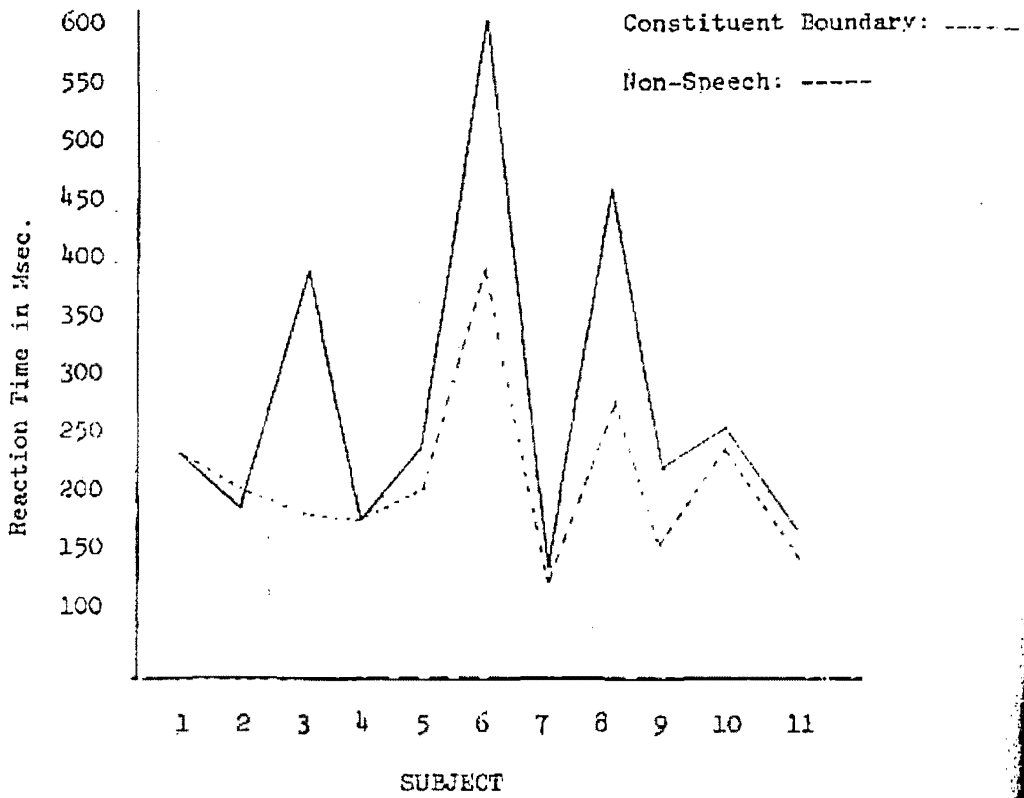


Fig. 12. Simple reaction time to click, and reaction time to click in a constituent boundary.

There is considerable variation in reaction time between subjects: subject 6, particularly, has quite slow reaction time to all conditions. Nevertheless, for each subject, the reaction times are in the same relationships, depending on the location of the click.

Click localization: The results of the click localization test are, in

general, in agreement with previous studies. Click localization tends to be accurate when the click occurs in a constituent boundary. This is shown in Fig. 13. The asterisk indicates the location of the click; the bar graph indicates the subjects' localization of the click.

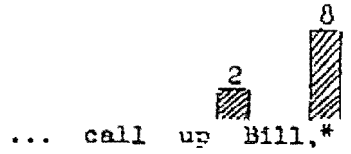
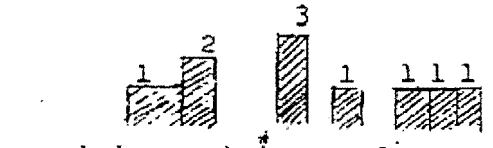
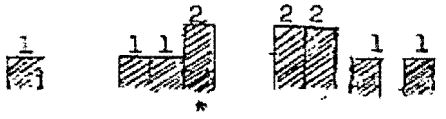


Fig. 13. Click localization when the click occurs in a constituent boundary.

There is also a tendency for subjects to move clicks towards deep structure constituent boundaries and to locate clicks between words. These results are shown in Fig. 14, for some typical sentences.





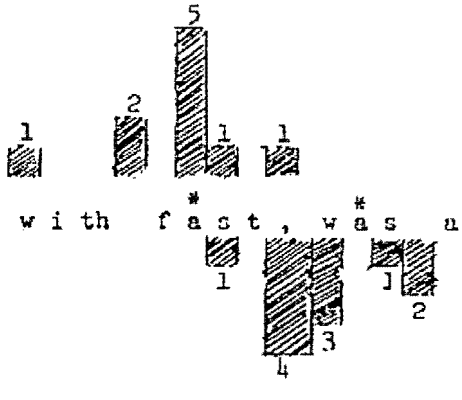


Fig. 14--continued

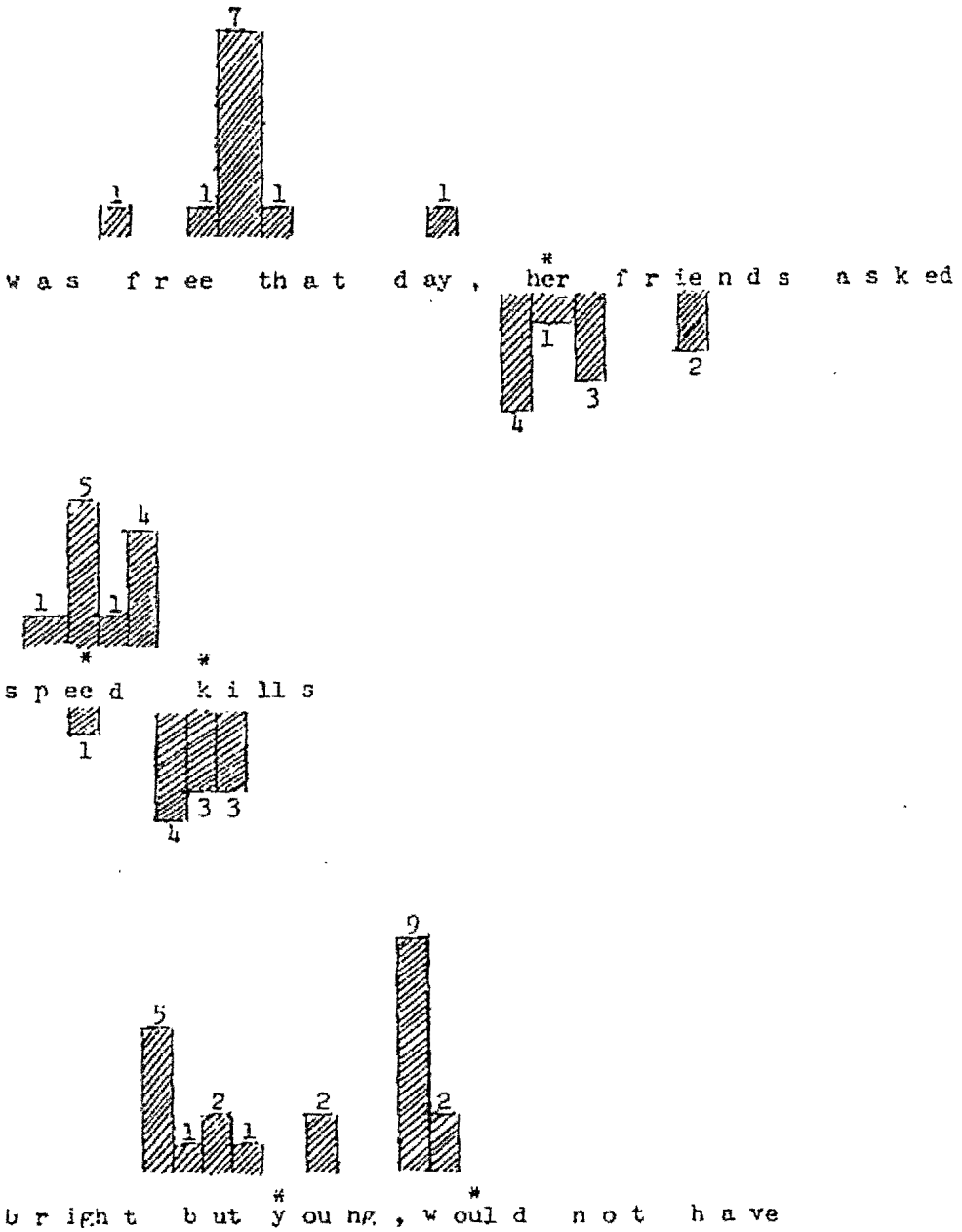


Fig. 14--continued

Fig. 14. Click localization.

However, the location of stress also affects click localization. Clicks in stressed vowels are localized much more accurately than clicks in unstressed vowels. This can be clearly seen by examining Fig. 15. The click in the stressed vowel of sleep is localized correctly more often than the click in the unstressed vowel of was. Furthermore, subjects do not miss the correct location by as much for the click in the stressed vowel as for the click in the unstressed vowel.

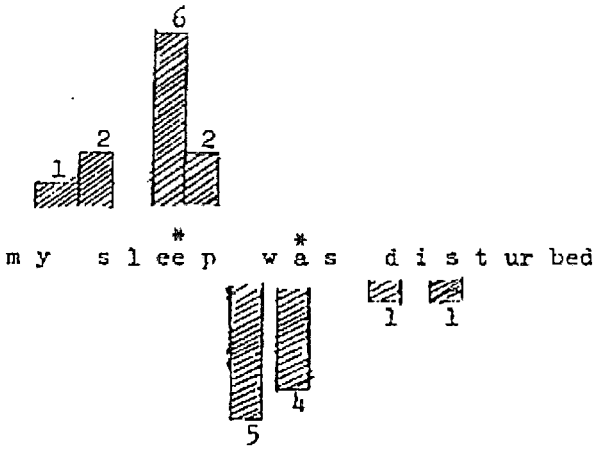


Fig. 15. Click localization in stressed and unstressed vowels.

Accuracy of click localization is summarized in Table 55.

TABLE 55

CLICK LOCALIZATION: PER CENT CORRECT

Stressed vowel	Unstressed vowel	Consonant	Constituent boundary
46	12	19	81

Discussion

The click localization data seem to imply that click localization is controlled by two parameters, constituent structure and the presence of stress. Click localization errors tend to lie in the direction predicted by theory, but clicks are less likely to be moved from a stressed vowel than from an unstressed vowel. That localization of clicks in consonants is also inaccurate may simply be a result of response bias: subjects may be less inclined to locate a click in a consonant. However, it may also result from the fact that the duration of consonants is short in relation to the duration of clicks.

The observed differences in reaction time imply that suprasegmental structure has some function in defining the units of speech perception. Since reaction time is not directly affected by constituent structure, it can be inferred that constituent structure does not define the units of perceptual input. Instead, the data support the hypothesis that units of perceptual input are defined by suprasegmental structure, i.e. stress and intonation.

There is one objection that might be raised to this conclusion. Stressed vowels occur in words that have semantic content whereas unstressed vowels occur in words that have less semantic content. In other words, words with stressed vowels are not predictable from context while words with unstressed vowels are much more readily predictable. The experiment, as designed, does not explicitly differentiate between this effect and the presence of stress. However, the objection is not crucial because the effect on reaction time is quite as pronounced when the click is in the consonant preceding the

vowel. It is difficult to see why a subject should react differently to these clicks if only the predictability of the word were the issue. Further testing is necessary, however, to rule out the "predictability hypothesis" completely.

CHAPTER FIVE

CONCLUSION

The results of the studies reported above are interesting in themselves, but they are also interesting in what they imply about the processes underlying speech perception. To summarize briefly, the results are the following:

1. Subjects are aware of sub-phonemic phonetic differences, at least under appropriate conditions, but can not make linguistic use of them.

2. Perception of at least some phonological segments involves special perceptual mechanisms, rather than proceeding segment-by-segment.

3. Syntactic units in perception may be defined by suprasegmental structure.

The Need for Perceptual Units

Before the implications of these findings for specific theories of speech perception will be discussed, it seems reasonable to re-examine the assumption of this study, namely that there are units in speech perception.

As Experiment I shows, subjects can become aware of very fine phonetic differences if they attend to a particular utterance with

great care. It is likely that subjects could even be taught to identify most of the words used in Experiment I properly, provided that the subjects got proper feedback, and provided that the stimuli were properly selected so that the distinctive cues were invariably present in each production. In this sense, there is no clear lower limit below which speech stimuli are perceived as "the same," and, one might suppose, no lower limit for a phonological perceptual unit either.

However, just because a listener can utilize fine phonetic detail when the conditions of a test force him to do so, does not imply that listeners inevitably notice or pay attention to such information. Rather, listeners are probably content with less detailed phonetic representations. To draw an analogy with visual perception, we do not examine leaves when we are looking at a forest. In visual perception, we can examine, in great detail, the shape and color of particular objects. But ordinarily, we do not do this; we are content to recognize objects and to behave appropriately to them--we sit in chairs, pat dogs, speak to our friends. Similarly, in the ordinary course of language use, we deal with something other than with fine phonetic differences. Therefore, there must be postulated some larger unit--or higher level--at which the phonological structure of an utterance is represented, independently of the fine phonetic details of the utterance.

This level, however, must be independent of syntactic or contextual information for the reason that new words, such as proper names and technical terms, do not present undue difficulty to us; we simply

hear the word, and we remember it.

These two considerations imply a lower and an upper boundary for the perception and coding of phonological information: the units involved in this process can not be equivalent to the phonetic representation of the utterance and the units can not be dependent on syntactic information.

Similarly, there must be some unit, or preferred units, in arriving at a syntactic analysis of a sentence. It is not possible for listeners to store a whole sentence in memory, simply because, unless the sentence were recoded in some way, it would very easily exceed the short-term memory capacity of a listener. It seems reasonable to suppose that the recoding operation can not process the sentence continuously as it is heard, but that the sentence must be broken up into some sort of units--perceptual segmentation units--for the recoding process to operate upon. The results of the recoding process certainly embody syntactic structure in some way.

It has been supposed previously that the perceptual segmentation units were syntactic as well. But the results of Experiment III can not be reconciled with the idea that segmentation units are syntactic. If they were, then reaction time to clicks and click localization should give the same results. Since this is not the case, the implication is that, at some level, sentences are processed in terms of non-syntactic units. The results of Experiment III imply that these units are defined by the phonological structure of an utterance and that these units function at the initial segmentation of the sentence. These initially segmented units are then recoded, probably by assigning them a particular syntactic function.

Thus, there is a need for at least two types of units in speech perception: units of phonological processing and units defining a part of a sentence for further syntactic analysis--perceptual segmentation units.

Implications for Perception Models

Not all of the theories of speech perception discussed in Chapter I make specific predictions about units of speech perception, but several do, namely the motor theory, analysis-by-synthesis, "filtering" theories, Osgood's perception model, and the perceptual strategies model. The experimental findings, reported above, conflict with some predictions made by these models, although, of course, the models may be revised slightly to cope with them.

First, the motor theory of speech perception, in that it asserts categorial perception of phonemes, conflicts with a listener's ability to become aware of sub-phonemic phonetic differences. If the perception of phonemes were indeed categorial, then listeners could not become aware of any sub-phonemic information whatever. Yet this is not the case; listeners are aware of sub-phonemic detail and use both vowel length and consonant quality in developing a strategy for making identification judgments. Second, that the motor theory postulates a phoneme-like unit as the basic unit of perception, it conflicts with the implications of Experiment II--that listeners apparently employ special perceptual mechanisms to process some consonant clusters, rather than perceiving the clusters "phoneme-by-phoneme."

This second objection also applies to analysis-by-synthesis models. These models assume that phonology is perceived in terms of

discrete segments. This assumption can not account for the finding of Experiment II--that reversal of the order of segments is the most common perceptual error.

In a fundamental way, the motor theory and analysis-by-synthesis are quite similar: both postulate that the listener generates a possible phonetic output and matches this output against the incoming message. The theories differ only in the nature of the internal mechanisms that they postulate. The experiments reported in this work do not have any implications for this basic postulate. However, it must be added here that there is no evidence that such internal mechanisms are strictly necessary. The "synthesis" theories have been postulated, apparently, because there are no invariants given immediately in the acoustic speech signal. Instead, the relationship between the acoustic signal and the perceptual result is quite complex.

Still, this difficulty is not unique to speech perception. In the study of visual perception, it has been commonly observed that the retinal image--which we may consider to be analogous to the acoustic input--is much more varied than the perception of objects. The retinal image changes radically as we view an object from different angles and from different distances, yet the percept is of an unchanging, stable object. The relationship between the retinal image and the percept is no less complex than the relationship between the acoustic signal and perceived speech, yet we do not posit a "motor theory of visual perception" for this reason.

These comments are added only to point out that a complex relationship is not sufficient grounds for positing intermediate devices of an unrelated type: theoretical mechanisms have to have

independent empirical justification.

The filtering theories discussed in Chapter I are of two types: theories that assume a phoneme-like unit, and Wickelgren's context-sensitive coding which assumes that the perceptual unit is similar to the traditional allophone. There are two objections to the phoneme-like unit: first, the well-known lack of invariance between phonemes and the acoustic signal and, second, the fact that obstruent clusters are apparently not perceived "phoneme-by-phoneme."

Wickelgren's theory tries to overcome the first difficulty by assuming smaller, hence presumably invariant, units, but it does so at the cost of proliferating the number of different units that must be assumed. Furthermore, it is still to be determined if there are invariar acoustic differences that can be used to determine the order of segments. Context-sensitive coding can, however, account for the perception of obstruent clusters. One further advantage of both types of filtering theories must be mentioned. Neither version of the theories is limited to a strict sequence of segments in the input, if the "filters" can be assumed to be working in parallel. Rather, the listener can be presumed to process a rather large segment of speech at one time.

Osgood argues that the word is the basic perceptual unit. However there are several difficulties with this position. First, as has already been pointed out, there must be some perceptual units which enable a listener to code a new word. It would be unparsimonious to suppose that these mechanisms are used only to code new words. Second, listeners can become aware of very subtle phonetic differences a finding which is counter to the notion that a word is the only

perceptual unit. But it seems likely that words function as units at some level of speech perception.

The perception of syntactic structure has been touched on only briefly in this study. The perceptual strategies suggested by Bever, and others, are not in dispute here; a fair amount of evidence has been offered to substantiate them, and no finding presented in this work conflicts with them. What has been questioned is the assumption that syntactic units provide the initial segmentation of a sentence. As has already been pointed out, this can not be the case because reaction time and click localization do not give the same results. Rather, the most likely hypothesis is that initial segmentation is accomplished by using the suprasegmental structure of an utterance. After this initial segmentation, perceptual strategies, as defined by Bever, may well apply to enable the listener to arrive at a syntactic analysis of the utterance.

The remarkable fact about speech perception is that it seems to be an easy and effortless process. Yet the mechanisms underlying this process are only beginning to be studied. Perhaps the best that could be said is that we are beginning to appreciate how complicated and mysterious the process of speech perception really is. Any adequate explanation will undoubtedly require a much more thorough understanding of human cognitive abilities on the one hand, and of the nature of language on the other.

BIBLIOGRAPHY

Abrams, K., and T. G. Bever, "Syntactic Structure Modifies Attention during Speech Perception and Recognition," Quarterly Journal of Experimental Psychology 21 (1969), 280-290.

Abramson, A. S., "Identification and Discrimination of Phonemic Tones," Journal of the Acoustical Society of America 33 (1961),8 842.

Abramson, A. S., and L. Lisker, "Discriminability along the Voicing Continuum: Cross-language Tests," Proceedings of the 6th International Congress of Phonetic Sciences, Prague (1967).

Bastian, J., and A. S. Abramson, "Identification and Discrimination of Phonemic Vowel Duration," Journal of the Acoustical Society of America 34 (1962), 743-644.

Bastian, J., P. Delattre, and A. M. Liberman, "Silent Interval as a Cue for the Distinction between Stops and Semivowels in Medial Position," Journal of the Acoustical Society of America 31 (1959), 1568.

Bastian, J., P. D. Eimas, and A. M. Liberman, "Identification and Discrimination of a Phonemic Contrast Induced by Silent Interval," Journal of the Acoustical Society of America 33 (1961), 842.

Bever, T. G., "The Cognitive Basis for Linguistic Structures," in Hayes, J., (ed.), Cognition and the Development of Language , New York: Wiley (1970).

Bever, T. G., J. Lackner, and R. Kirk, "The Underlying Structure Sentence is the Primary Unit of Speech Perception," Perception and Psychophysics 5 (1969), 225-234.

Bever, T. G. and D. T. Langendoen, "The Interaction of Speech Perception and Grammatical Structure in the Evolution of Language," unpublished manuscript.

Blesser, B. A., "Inadequacy of a Spectral Description in Relationship to Speech Perception," Paper presented at the 78th meeting of the Acoustical Society of America (4-7 November 1969).

Bloomfield, Leonard, Language, New York (1933).

Blumenthal, A., "Prompted Recall of Sentences," Journal of Verbal Learning and Verbal Behavior 6 (1967), 203-206.

Bondarko, L. V., N. G. Zagorujko, V. A. Koževnikov, A. P. Molčanov, and L. A. Čistovič, "A Model of Speech Perception by Humans," (Ilse Lehiste, translator), Working Papers in Linguistics 6 (The Ohio State University, 1970), 89-132.

Broadbent, D. E., and M. Gregory, "Accuracy of Recognition for Speech Presented to the Right and Left Ears," Quarterly Journal of Experimental Psychology 16 (1964), 359-60.

Broadbent, D. E., and P. Ladefoged, "Auditory Perception of Temporal Order," Journal of the Acoustical Society of America 31 (1959), 1539.

Bryden, M. P., "Ear Preferences in Auditory Perception," Journal of Experimental Psychology 65 (1963), 103-105.

Clark, H. H. and E. V. Clark, "Semantic Distinctions and Memory for Complex Sentences," Quarterly Journal of Experimental Psychology 20 (1968), 129-138.

Clark, H. H. and R. A. Stafford, "Memory for Semantic Features," Journal of Experimental Psychology 80 (1969), 326-334.

Cooper, F. S., "Research on Reading Machines for the Blind," in P. A. Zahl (ed.), Blindness: Modern Approaches to the Unseen Environment, Princeton Univ. Press (1950).

Cooper, F. S., "Describing the Speech Process in Motor Command Terms," Journal of the Acoustical Society of America 39 (1966), 1221 A.

Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst, and H. G. Gerstman, "Some Experiments on the Perception of Synthetic Speech Sounds," Journal of the Acoustical Society of America 24 (1952), 597-606.

Cooper, F. S., A. M. Liberman, and J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," Proceedings of the National Academy of Sciences 37 (1951), 318-325.

Cross, D. V. and H. L. Lane, "On the Discriminative Control of Concurrent Responses: The Relations among Response Frequency, Latency, and Topography in Auditory Generalization," Journal of the Experimental Analysis of Behavior 5 (1962), 487-496.

Cross, D. V., H. L. Lane, and W. C. Sheppard, "Identification and Discrimination Functions for a Visual Continuum and Their Relation to the Motor Theory of Speech Perception," Journal of Experimental Psychology 70 (1965), 63-74.

Davis, H., "Auditory Communication," Journal of Speech and Hearing Disorders 16 (1951), 3-8.

Delattre, P. C., A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," Journal of the Acoustical Society of America 27 (1955), 679-773.

Delattre, P., A. M. Liberman, F. S. Cooper, and L. G. Gerstman, "An Experimental Study of the Acoustic Determinants of Vowel Color," Word 8 (1952), 195-210.

Denes, P., "Effect of Duration on the Perception of Voicing," Journal of the Acoustical Society of America 27 (1955), 761-766.

Denes, P., "On the Motor Theory of Speech Perception," Proceedings of the 5th International Congress of Phonetic Sciences, (Münster, Basel, New York: S. Karger, 1964) 232-238.

Dixon, Theodore R., and David L. Horton (eds.), Verbal Behavior and General Behavior Theory (Englewood Cliffs., N.J.: Prentice-Hall, 1968).

Eimas, P., "The Relation between Identification and Discrimination along Speech and Non-speech Continua," Language and Speech 6 (1963), 206-217.

Fairbanks, G., "A Theory of the Speech-Mechanism as a Servo-system," Journal of Speech and Hearing Disorders 19 (1954), 133-139.

Flanagan, J. H., "Difference Limen for Vowel Formant Frequency," Journal of the Acoustical Society of America 27 (1955), 613-617.

Fodor, J. A., and T. G. Bever, "The Psychological Reality of Linguistic Segments," Journal of Verbal Learning and Verbal Behavior 4 (1965), 414-420.

Fodor, J., and M. Garrett, "Some Syntactic Determinants of Sentential Complexity," Perception and Psychophysics 2 (1967), 289-296.

Fodor, J., M. Garrett, and T. Bever, "Some Syntactic Determinants of Sentential Complexity II: Verb Structure," Perception and Psychophysics 3 (1968), 453-461.

Fromkin, V. A., "Neuro-muscular Specification of Linguistic Units," Language and Speech 9 (1966), 170-199.

Fry, D. B., "Duration and Intensity as Physical Correlates of Linguistic Stress," Journal of the Acoustical Society of America 27 (1955), 765-768.

Fry, D. B., "Perception and Recognition in Speech," in For Roman Jakobson, M. Halle (ed.), (The Hague: Mouton & Co., 1956).

Fry, D. B., "Experimental Evidence for the Phoneme," in In Honor of Daniel Jones, D. Abercrombie, et al. (eds.), (London: Longmans, 1964).

Fry, D. B., "The Function of the Syllable," Zeitschrift für Phonetik 17 (1964), 215-221.

Fry, D. B., "Reaction Time Experiments in the Study of Speech Processing," Progress Report, Phonetics Laboratory, University College, London (1968).

Fry, D. B., A. S. Abramson, P. D. Eimas, and A. M. Liberman, "The Identification and Discrimination of Synthetic Vowels," Language and Speech 5 (1962), 171-189.

Fry, D. B. and P. Denes, "An Analogue of the Speech Recognition Process," Mechanisation of Thought (London, 1958).

Garrett, M., T. G. Bever, and J. A. Fodor, "The Active Use of Grammar in Speech Perception," Perception and Psychophysics 1 (1966), 30-32.

Gibson, Eleanor J., Principles of Perceptual Learning and Development. (Appleton-Century-Crofts: New York, 1969).

Gibson, James J., The Senses Considered as Perceptual Systems, (Houghton Mifflin: Boston, 1966).

Greenberg, J. H., and J. J. Jenkins, "Studies in the Psychological Correlates of the Sound System of American English," Word (1964), 157-177.

Halle, M., G. W. Hughes, and J.-P. A. Radley, "Acoustic Properties of Stop Consonants," Journal of the Acoustical Society of America 29 (1957), 107-116.

Halle, M., and K. N. Stevens, "Speech Recognition: A Model and a Program for Research," in The Structure of Language: Readings in the Philosophy of Language, J. A. Fodor and J. J. Katz, (eds.), (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

Harris, K., "Cues for the Discrimination of American English Fricatives in Spoken Syllables," Language and Speech 1 (1958), 1-7.

Harris, K., J. Bastian, and A. M. Liberman, "Mimicry and the Perception of a Phonemic Contrast Induced by Silent Interval: Electromyographic and Acoustic Measures," Journal of the Acoustical Society of America 33 (1961), 842.A.

Harris, K., H. Hoffman, A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Effect of Third-formant Transitions in the Perception of the Voiced Stop Consonants," Journal of the Acoustical Society of America 30 (1958), 122-126.

Hirsch, I. J., "Auditory Perception of Temporal Order," Journal of the Acoustical Society of America 31 (1959), 759-767.

Hockett, C., A Manual of Phonology, Indiana University Publications in Anthropology 17 (1955).

Hoffman, Howard S., "A Study of Some Cues in the Perception of the Voiced Stop Consonants," Journal of the Acoustical Society of America 30 (1958), 1035-1041.

House, A. S. and G. Fairbanks, "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," Journal of the Acoustical Society of America 25 (1953), 105-113.

House, A. S., K. N. Stevens, T. Sandel, and G. Arnold, "On the Learning of Speech-like Vocabularies," Journal of Verbal Learning and Verbal Behavior 1 (1962), 133-143.

Hughes, G. W. and M. Halle, "Spectral Properties of Fricative Consonants," Journal of the Acoustical Society of America 28 (1956), 303-310.

Jakobovits, L. A., and M. S. Miron (eds.), Readings in the Psychology of Language, (Englewood Cliffs, N.J.: Prentice-Hall, 1967).

Johnson, N., "The Psychological Reality of Phrase-structure Rules," Journal of Verbal Learning and Verbal Behavior 5 (1966), 469-475.

Joos, M., Acoustic Phonetics, Supplement to Language 24 (1948).

Katz, G., "Mentalism in Linguistics," Language 40 (1964), 124-138.

Kozhevnikov, V. A. and L. A. Chistovich, Speech: Articulation and Perception. U. S. Department of Commerce, JPRS Report 30,543 (1965).

Ladefoged, P., "The Perception of Speech," in Mechanisation of Thought Processes, VI (London, 1959).

Ladefoged, P., Three Areas of Experimental Phonetics, (London: Oxford University Press, 1967).

Ladefoged, P., and D. E. Broadbent, "Information Conveyed by Vowels," Journal of the Acoustical Society of America 29 (1957), 98.

Ladefoged, P., and D. E. Broadbent, "Perception of Sequence in Auditory Events," Quarterly Journal of Experimental Psychology 12 (1960) 162-70.

Lane, H. L., "Psychophysical Parameters of Vowel Perception," Psychological Monographs 76 (1962), 44.

Lane, H. L., "The Motor Theory of Speech Perception: A Critical Review," Psychological Review 72 (1965), 275-309.

Lehiste, I., Suprasegmentals (Cambridge, Mass.: M.I.T. Press, 1970).

Lehiste, I., "The Temporal Realization of Morphological and Syntactic Boundaries," Paper presented at the 81st Meeting of the Acoustical Society of America (April, 1971).

Lehiste, I. and G. E. Peterson, "Vowel Amplitude and Phonemic Stress in American English," Journal of the Acoustical Society of America 31 (1959), 428-435.

Liberman, A. M., "Some Results of Research on Speech Perception," Journal of the Acoustical Society of America 29 (1957), 117-123.

Liberman, A. M., F. S. Cooper, K. S. Harris, and P. F. MacNeilage, "A Motor Theory of Speech Perception," Proceedings of the Speech Communication Seminar (Stockholm: Royal Institute of Technology, 1962).

Liberman, A. M., F. S. Cooper, K. S. Harris, P. F. MacNeilage, and M. Studdert-Kennedy, "Some Observations on a Model for Speech Perception," Models for the Perception of Speech and Visual Form (M.I.T. Press, 1965).

Liberman, A. M., F. S. Cooper, D. S. Shankweiler, and M. Studdert-Kennedy, "Perception of the Speech Code," Psychological Review 74 (1967), 431-461.

Liberman, A. M., P. C. Delattre, and F. S. Cooper, "The Role of Selected Stimulus Variables in the Perception of the Unvoiced Stop Consonants," American Journal of Psychology 65 (1952), 497-516.

Liberman, A. M., P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Positions," Language and Speech 1 (1958), 153-167.

Liberman, A. M., P. C. Delattre, F. S. Cooper, and H. J. Gerstman, "The Role of Consonant-vowel Transitions in the Perception of Stop and Nasal Consonants," Psychological Monograph 68 (1954).

Liberman, A. M., P. C. Delattre, H. J. Gerstman, and F. S. Cooper, "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," Journal of Experimental Psychology 52 (1956), 127-137.

Liberman, A. M., K. S. Harris, P. Eimas, L. Lisker, and J. Bastian, "An Effect of Learning on Speech Perception: The Discrimination of Durations of Silence with and without Phonemic Significance," Language and Speech 4 (1961), 175-195.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," Journal of Experimental Psychology 54 (1957), 358-367.

Liberman, A. M., K. S. Harris, J. Kinney, and H. L. Lane, "The Discrimination of Relative Onset Time of the Components of Certain Speech and Non-speech Patterns," Journal of Experimental Psychology 61 (1961), 379-388.

Licklider, J. C. R., "On the Process of Speech Perception," Journal of the Acoustical Society of America 24 (1952), 590-594.

Licklider, J. C. R., and G. A. Miller, "The Perception of Speech," in Handbook of Experimental Psychology, S. S. Stevens, ed., (New York: John Wiley & Sons, 1951).

Lieberman, P., "Some Effects of Semantic and Grammatical Context on the Production and Perception of Speech," Language and Speech 6 (1963), 172-187.

Lieberman, P., Intonation, Perception, and Language (Cambridge, Mass.: M.I.T. Press, 1967).

Lindgren, N., "Machine Recognition of Human Language," IEEE Spectrum (March, 1965), 114-136; (April, 1965), 45-59.

Lisker, L., "Minimal Cues for Separating /w,r,l,j/ in Intervocalic Position," Word 13 (1957), 257-267.

Lisker, L., "Closure Duration and the Intervocalic Voiced/Voiceless Distinction in English," Language 33 (1957), 42-49.

Lisker, L., "Anatomy of Unstressed Syllables," Journal of the Acoustical Society of America 30 (1958), 682, A.

Lisker, L., and A. S. Abramson, "Cross-language Study of Voicing in Initial Stops: Acoustical Measurements," Word 20 (1964), 384-422.

Lisker, L., and A. S. Abramson, "Stop Categories and Voice Onset Time," in Proceedings of the Fifth International Congress of Phonetic Sciences, (Münster, 1964. Basel: S. Karger, 1965).

Lisker, L., and A. S. Abramson, "The Voicing Dimension: Some Experiments in Comparative Phonetics," in Proceedings of the Sixth International Congress of Phonetic Sciences, (Prague, 1967).

Lotz, J., A. S. Abramson, H. Gerstman, F. Ingemann, and W. J. Nemser, "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai," Language and Speech 3 (1960), 71-77.

Lisker, L., F. S. Cooper, and A. M. Liberman, "The Uses of Experiment in Language Description," Word 18 (1962), 82-106.

Lyons, J., and R. G. Wales (eds.), Psycholinguistics Papers (Edinburgh: Edinburgh University Press, 1966).

Malécot, A., "Acoustic Cues for Nasal Consonants," Language 32 (1956), 274-284.

Mattingly, I., and A. M. Liberman, "The Speech Code and the Physiology of Language," in K. N. Leibovic, Information Processing in the Nervous System (New York: Springer-Verlag, 1969).

Mehler, J., "Some Effects of Grammatical Transformations on the Recall of English Sentences," Journal of Verbal Learning and Verbal Behavior 2 (1963), 346-351.

Mehler, J., and P. Carey, "Role of Surface and Base Structure in the Perception of Sentences," Journal of Verbal Learning and Verbal Behavior 6 (1967), 335-338.

Miller, G. A., "The Perception of Short Bursts of Noise," Journal of the Acoustical Society of America 20 (1948), 160-170.

Miller, G. A., "Speech and Language," in Handbook of Experimental Psychology, S. S. Stevens, ed., (New York: John Wiley & Sons, 1951).

Miller, G. A., "The Magical Number 7, Plus or Minus 2: Some Limits in Our Capacity for Processing Information," Psychological Review 63 (1950), 81-97.

Miller, G. A., "The Perception of Speech," in For Roman Jakobson: Essays on the Occasion of His 60th Birthday, M. Halle, ed., (The Hague: Mouton & Co., 1956).

Miller, G. A., "Speech and Communication," Journal of the Acoustical Society of America 30 (1958), 397-398.

Miller, G. A., "Some Psychological Studies of Grammar," American Psychologist 17 (1962), 748-762.

Miller, G. A., "Decision Units in the Perception of Speech," IRE Transactions on Information Theory IT-8 (1962), 81-83.

Miller, G. A., E. Galanter, and K. H. Pribram, Plans and the Structure of Behavior (New York: Henry Holt & Co., 1960).

Miller, G. A., G. A. Heise, and W. Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials," Journal of Experimental Psychology 41 (1951), 329-335.

Miller, G. A., and S. Isard, "Some Perceptual Consequences of Linguistic Rules," Journal of Verbal Learning and Verbal Behavior 2 (1963), 217-228.

Miller, G. A., and P. E. Nicely, "An Analysis of Perceptual Confusions among Some English Consonants," Journal of the Acoustical Society of America 27 (1955), 338-352.

Mowrer, O. H., "The Psychologist Looks at Language," American Psychologist 9 (1954), 660-694.

Norman, D. A., Memory and Attention: An Introduction to Human Information Processing (New York: Wiley, 1969).

O'Connor, J. D., H. J. Gerstman, A. M. Liberman, P. S. Delattre, and F. S. Cooper, "Acoustic Cues for the Perception of Initial /w,j,r,l/ in English," Word 13 (1957), 24-43.

Öhman, S. E. G., "Coarticulation in VCV Utterances: Spectrographic Measurements," Journal of the Acoustical Society of America 39 (1966), 151-168.

Osgood, Charles E., "Psycholinguistics," in Psychology: A Study of a Science, S. Koch, ed. (New York: McGraw-Hill, 1963).

Osgood, Charles E., "On Understanding and Creating Sentences," American Psychologist 18 (1963), 735-751.

Peterson, G. E., "The Phonetic Value of Vowels," Language 27 (1951), 541-553.

Peterson, G. E., "The Information-bearing Elements of Speech," Journal of the Acoustical Society of America 24 (1952), 624-637.

Peterson, G. E., "Basic Physical Systems for Communication between Two Individuals," Journal of Speech and Hearing Disorders 18 (1953), 116-120.

Peterson, G. E., "An Oral Communication Model," Language 31 (1955), 414-427.

Peterson, G. E., and H. L. Barney, "Control Methods Used in a Study of the Vowels," Journal of the Acoustical Society of America 24 (1952), 175-184.

Schatz, C. D., "The Role of Context in the Perception of Stops," Language 30 (1954), 47-56.

Shankweiler, D., and M. Studdert-Kennedy, "An Analysis of Perceptual Confusions in Identification of Dichotically Presented CVC Syllables," Journal of the Acoustical Society of America 41 (1967), 1581.

Shankweiler, D., and M. Studdert-Kennedy, "Identification of Consonants and Vowels Presented to Left and Right Ears," Quarterly Journal of Experimental Psychology 19 (1967), 59-63.

Skinner, B. F., Verbal Behavior (New York: Appleton-Century-Crofts, 1957).

Stevens, K. N., "The Perception of Vowel Formants," Journal of the Acoustical Society of America 24 (1952), 450.

Stevens, K. N., "Toward a Model for Speech Recognition," Journal of the Acoustical Society of America 32 (1960), 47-55.

Stevens, K. N., and M. Halle, "Remarks on Analysis by Synthesis and Distinctive Features," Models for the Perception of Speech and Visual Form (Cambridge, Mass.: M.I.T. Press, 1965).

Stevens, K. N., A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Öhman, "Cross-language Study of Vowel Perception," Language and Speech 12 (1969), 1-23.

Stevens, K. N., S. E. G. Öhman, and A. M. Liberman, "Identification and Discrimination of Rounded and Unrounded Vowels," Journal of the Acoustical Society of America 35 (1963), 1900.A.

Stolz, W. S., "A Study of the Ability to Decode Grammatically Novel Sentences," Journal of Verbal Learning and Verbal Behavior 6 (1967), 867-873.

Strevens, P., "Spectra of Fricative Noise in Human Speech," Language and Speech 3 (1960), 32-48.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper, "Motor Theory of Speech Perception: A Reply to Lane's Critical Review," Psychological Review 77 (1970), 234-249.

Studdert-Kennedy, M., A. M. Liberman, and K. N. Stevens, "Reaction Time to Synthetic Stop Consonants and Vowels at Phoneme Centers and at Phoneme Boundaries," Journal of the Acoustical Society of America 35 (1963), 1900. A.

Uldall, E., "Transitions in Fricative Noise," Language and Speech 7 (1964), 13-15.

Warren, R. M., and R. P. Warren, "Auditory Illusions and Confusions," Scientific American (December, 1970), 30-36.

Watson, John B., Behaviorism (New York: W. W. Norton, 1930).

Wickelgren, Wayne A., "Distinctive Features and Errors in Short-Term Memory for English Consonants," Journal of the Acoustical Society of America 39 (1966), 388-398.

Wickelgren, W. A., "Context-sensitive Coding, Associative Memory, and Serial Order in (Speech) Behavior," Psychological Review 76 (1969a), 1-15.

Wickelgren, W. A., "Context-sensitive Coding in Speech Recognition, Articulation, and Development," in Information Processing in the Nervous System, K. N. Leibovic, ed. (New York: Springer-Verlag, 1969b).

Wright, C. W. An English Word Count (Pretoria: National Bureau of Educational and Social Research, 1965).